

*Руководство по
администрированию и
производительности IBM SPSS
Modeler Server 16*



Примечание

Прежде чем использовать эту информацию и продукт, описанный в ней, прочтите сведения в разделе “Уведомления” на стр. 71.

Информация о продукте

Это издание применимо к версии 16, выпуску 0, модификации 0 IBM(r) SPSS(r) и ко всем последующим выпускам и модификациям до тех пор, пока в новых изданиях не будет указано иное.

Содержание

Предисловие	v
-----------------------	---

Глава 1. О программе IBM SPSS Modeler 1

Продукты IBM SPSS Modeler	1
IBM SPSS Modeler	1
сервер IBM SPSS Modeler	1
IBM SPSS Modeler Administration Console	2
IBM SPSS Modeler Batch	2
IBM SPSS Modeler Solution Publisher	2
Адаптеры сервер IBM SPSS Modeler для IBM SPSS Collaboration and Deployment Services	2
Выпуски IBM SPSS Modeler	2
Документация IBM SPSS Modeler	3
Документация SPSS Modeler Professional	3
Документация SPSS Modeler Premium	4
Примеры прикладных программ	4
Папка demos	5

Глава 2. Рекомендации по архитектуре и аппаратным средствам 7

Архитектура IBM SPSS Modeler	7
Описание архитектуры	7
Рекомендации по аппаратным средствам	9
Временное дисковое пространство и требования к оперативной памяти	10
Доступ к данным	11
Ссылки на файлы данных	12
Импорт файлов данных IBM SPSS Statistics	12
Инструкции по установке	13

Глава 3. Поддержка IBM SPSS Modeler 15

Соединение с сервер IBM SPSS Modeler	15
Конфигурирование единой регистрации	16
Добавление и изменение соединений с сервер IBM SPSS Modeler	17
Поиск серверов в IBM SPSS Collaboration and Deployment Services	18
Данные и файловые системы	18
Аутентификация пользователей	19
Создание файлов	19
Различия в результатах	19

Глава 4. Администрирование сервера IBM SPSS Modeler 21

Запуск и остановка сервера IBM SPSS Modeler	21
Для запуска, остановки и проверки состояния в Windows	21
Для запуска, остановки и проверки состояния в UNIX	21
Обработка неотвечающих процессов сервера (системы UNIX)	22
Администрирование	22

Использование IBM SPSS Modeler Administration Console	23
Запуск Modeler Administration Console	23
Конфигурирование доступа с Modeler Administration Console	23
Соединения сервер SPSS Modeler	23
Конфигурирование сервера IBM SPSS Modeler	24
Мониторинг сервер SPSS Modeler	29
Использование файла options.cfg	29
Закрытие неиспользуемых соединений с базой данных	29
Использование SSL для защищенной передачи данных	30
Как работает SSL	30
Обеспечение защищенной связи клиент-сервер и сервер-сервер с использованием SSL	30

Глава 5. Обзор производительности 35

Параметры производительности и оптимизации сервера	35
Параметры производительности и оптимизации клиента	35
Использование и оптимизация базы данных	37
Оптимизация SQL	37

Глава 6. Оптимизация SQL 39

Как работает генерирование SQL	40
Пример генерирования SQL	40
Конфигурирование оптимизации SQL	42
Предварительный просмотр сгенерированного SQL	43
Просмотр SQL для слепков моделей	43
Советы по максимизации генерирования SQL	43
Узлы, поддерживающие генерирование SQL	44
Выражения и операторы CLEM, поддерживающие генерирование SQL	48
Использование функций SQL в выражениях CLEM	51
Написание запросов SQL	51

Приложение А. Конфигурирование Oracle на платформах UNIX 53

Конфигурирование Oracle для оптимизации SQL	53
---	----

Приложение В. Конфигурирование сценариев запуска UNIX 55

Введение	55
Сценарии	55
Автоматический запуск и остановка сервер IBM SPSS Modeler	55
Запуск и остановка сервера IBM SPSS Modeler вручную	56
Редактирование сценариев	56
Управление разрешениями при создании файла	56
Сервер IBM SPSS Modeler и пакет Data Access Pack	56
Устранение неисправностей конфигурации ODBC	58

Пути библиотек	61
Конфигурирование драйвера для узла представления предприятия	61

**Приложение С. Конфигурирование и
запуск сервер SPSS Modeler в
качестве процесса обычного
пользователя (не root) в UNIX 63**

Введение	63
Конфигурирование в качестве обычного пользователя (не root) без помощи базы данных секретных паролей .	63
Конфигурирование в качестве рядового пользователя (не root) с помощью базы данных секретных паролей .	64
Запуск сервера SPSS Modeler с учетной записью обычного пользователя (не root).	65
Устранение неисправностей при аутентификации пользователей	65

**Приложение D. Конфигурирование и
запуск сервера SPSS Modeler при
помощи файла секретных паролей в
Windows 67**

Введение	67
Конфигурирование с использованием базы данных секретных паролей	67

**Приложение E. Балансировка
нагрузок при помощи кластеров
серверов 69**

Уведомления. 71

Товарные знаки	72
--------------------------	----

Индекс 75

Предисловие

IBM® SPSS Modeler - это инструментальная среда исследования данных IBM Corp., рассчитанная на работу с предприятием. SPSS Modeler помогает организациям улучшать отношения с заказчиками и общественностью через всестороннее понимание данных. Организации используют понимание, полученное от SPSS Modeler, чтобы удержать выгодных заказчиков, обнаружить возможности сопутствующих продаж, привлечь новых заказчиков, обнаружить мошенничество, снизить риск и улучшить предоставление государственных услуг.

Визуальный интерфейс SPSS Modeler привлекателен для пользователей, опыт которых относится к своей отрасли, а не IT-технологиям. При их участии создаются модели с большей предсказательной силой и уменьшается срок разработки решения. SPSS Modeler предлагает целый ряд технологий моделирования - алгоритмы прогнозирования, классификации, сегментации и обнаружения ассоциаций. После создания моделей IBM SPSS Modeler Solution Publisher предлагает средства их распространения в масштабах предприятия среди сотрудников, принимающих решения, или внедрения в базу данных.

О бизнес аналитике IBM

Программное обеспечение IBM для бизнес аналитики предоставляет полную, последовательную и точную информацию, которая повышает эффективность ведения бизнеса. Полный набор программного обеспечения для business intelligence, прогностической аналитики, управления финансовой эффективностью и стратегией и аналитических приложений позволяет ясно видеть текущую ситуацию, а также делать прогнозы, позволяющие предпринимать практические действия. В сочетании с решениями для конкретных отраслей, проверенной практикой и услугами бизнес аналитика IBM позволяет организациям любых размеров достигать наивысшей производительности, уверенно автоматизировать процессы принятия решений и добиться лучших результатов.

Как составная часть этого набора, программное обеспечение IBM SPSS Predictive Analytics помогает организациям предсказывать будущие события и предпринимать практические действия непосредственно на основе этих предсказаний. Во всем мире коммерческие, государственные и научные организации полагаются на технологию IBM SPSS, которая дает конкурентное преимущество в привлечении, удержании и приросте заказчиков при сокращении возможностей мошенничества и риска ущерба от него. Внедряя программное обеспечение IBM SPSS в повседневную деятельность, организации становятся прогнозирующими предприятиями - способными руководить и автоматизировать руководящие решения, чтобы достигать бизнес-целей и ощутимого конкурентного преимущества. Чтобы получить дальнейшую информацию или связаться с представителем, зайдите на <http://www.ibm.com/spss>.

Техническая поддержка

Техническая поддержка предоставляется клиентам, оплачивающим обновительные взносы. Пользователи могут обращаться в службу технической поддержки, если у них возникают какие-либо проблемы с использованием или установкой программного обеспечения IBM Corp.. За технической поддержкой обращайтесь на сайт IBM Corp. по адресу <http://www.ibm.com/support>. При обращении за поддержкой будьте готовы назвать себя и организацию, в которой вы работаете.

Глава 1. О программе IBM SPSS Modeler

IBM SPSS Modeler - это комплект инструментов исследования данных, при помощи которого можно быстро разрабатывать прогнозные модели, использующие деловые знания и опыт, и внедрять их в деловые операции для усовершенствования процесса принятия решений. Разработанный на основе модели промышленного стандарта CRISP-DM, IBM SPSS Modeler поддерживает весь процесс исследования данных, от обработки исходных данных до получения лучших деловых результатов.

IBM SPSS Modeler предлагает ряд методов моделирования, взятых из таких областей, как обучение машин, искусственный интеллект и статистика. При помощи методов, доступных на палитре Моделирование, можно извлечь новую информацию из данных и разработать прогнозные модели. У каждого из методов есть свои сильные стороны и типы задач, для решения которых он лучше всего подходит.

SPSS Modeler можно приобрести как отдельный продукт или использовать как клиент в сочетании с сервер SPSS Modeler. Кроме того, доступен ряд дополнительных возможностей, сводка которых дается в следующих разделах. Дополнительную информацию смотрите в разделе <http://www.ibm.com/software/analytics/spss/products/modeler/>.

Продукты IBM SPSS Modeler

В семейство продуктов IBM SPSS Modeler и связанные с этим семейством программы входят следующие продукты:

- IBM SPSS Modeler
- сервер IBM SPSS Modeler
- IBM SPSS Modeler Administration Console
- IBM SPSS Modeler Batch
- IBM SPSS Modeler Solution Publisher
- Адаптеры сервер IBM SPSS Modeler для IBM SPSS Collaboration and Deployment Services

IBM SPSS Modeler

SPSS Modeler - это полнофункциональная версия продукта, устанавливаемая и запускаемая на персональном компьютере. SPSS Modeler можно запустить в локальном режиме, как автономный продукт, или в распределенном режиме вместе с сервер IBM SPSS Modeler, чтобы повысить производительность на больших наборах данных.

Используя SPSS Modeler, можно быстро и интуитивно строить точные прогнозные модели, не прибегая к программированию. Используя уникальный визуальный интерфейс, можно легко визуализировать процесс анализа данных. В продукт встроены расширенные функции аналитики, при поддержке которых можно обнаруживать в данных скрытые структуры и тенденции. Можно моделировать результаты и выяснять, какие факторы на них влияют, чтобы полностью использовать деловые возможности и ограничивать риски.

SPSS Modeler доступен в двух версиях: SPSS Modeler Professional и SPSS Modeler Premium. Дополнительную информацию смотрите в разделе “Выпуски IBM SPSS Modeler” на стр. 2.

сервер IBM SPSS Modeler

SPSS Modeler пользуется архитектурой клиент - сервер, чтобы распределять требования ресурсоемких операций по мощным серверным программам, что повышает производительность для больших наборов данных.

сервер SPSS Modeler - это отдельно лицензируемый продукт, который непрерывно работает в режиме распределенного анализа на хосте сервера совместно с одной или несколькими установками IBM SPSS Modeler. При этом сервер SPSS Modeler обеспечивает высокую производительность для больших наборов данных, поскольку ресурсоемкие операции можно выполнять на сервере без скачивания данных на компьютер клиента. Кроме того, сервер IBM SPSS Modeler обеспечивает поддержку для возможностей оптимизации SQL и моделирования в базе данных, что дает дополнительный выигрыш в производительности и автоматизации.

IBM SPSS Modeler Administration Console

Modeler Administration Console - это графическая программа для управления многочисленными опциями конфигурации сервера SPSS Modeler, который также можно конфигурировать посредством файла опций. Эта прикладная программа содержит консольный пользовательский интерфейс для отслеживания и конфигурирования установок сервера SPSS Modeler installations, and is available free-of-charge сервера SPSS Modeler. Эту прикладную программу можно установить только на компьютерах Windows; однако она может управлять сервером на любой поддерживаемой платформе.

IBM SPSS Modeler Batch

Хотя обычно исследование данных - интерактивный процесс, можно также запустить SPSS Modeler из командной строки, не открывая графический интерфейс. Например, у вас могут быть продолжительные или повторяющиеся задачи, которые желательно выполнить без участия пользователя. SPSS Modeler Batch - это особая версия продукта, предоставляющая поддержку всех аналитических возможностей SPSS Modeler без вызова обычного пользовательского интерфейса. Для использования SPSS Modeler Batch требуется лицензия сервера SPSS Modeler.

IBM SPSS Modeler Solution Publisher

SPSS Modeler Solution Publisher - это инструмент, при помощи которого можно создать пакетную версию потока SPSS Modeler; такую версию можно запускать внешним механизмом времени выполнения или встроить во внешнюю прикладную программу. Этим способом можно публиковать и внедрять полные потоки SPSS Modeler для использования в средах, где SPSS Modeler не установлен. SPSS Modeler Solution Publisher распространяется в составе службы IBM SPSS Collaboration and Deployment Services - Scoring, для которой требуется отдельная лицензия. С этой лицензией вы получаете модуль времени выполнения SPSS Modeler Solution Publisher, при помощи которого можете запускать опубликованные потоки.

Адаптеры сервера IBM SPSS Modeler для IBM SPSS Collaboration and Deployment Services

Для IBM SPSS Collaboration and Deployment Services доступен ряд адаптеров, при помощи которых SPSS Modeler и сервер SPSS Modeler могут взаимодействовать с репозиторием IBM SPSS Collaboration and Deployment Services. При этом поток SPSS Modeler, внедренный в репозиторий, доступен для совместного использования несколькими пользователями или для обращения из прикладной программы IBM SPSS Modeler Advantage тонкого клиента. Адаптер устанавливается в той системе, в которой находится репозиторий.

Выпуски IBM SPSS Modeler

SPSS Modeler доступен в следующих выпусках.

SPSS Modeler Professional

SPSS Modeler Professional содержит все инструменты, необходимые для работы с большинством типов структурированных данных, таких как трассировка поведения и взаимодействия в системах CRM, демографии, поведения покупателей и данных о продажах.

SPSS Modeler Premium

SPSS Modeler Premium - это отдельно лицензируемый продукт, расширяющий SPSS Modeler Professional для работы с такими специальными данными, как данные в аналитике объектов или социальных сетях, и с неструктурированными текстовыми данными. SPSS Modeler Premium состоит из следующих компонентов.

IBM SPSS Modeler Entity Analytics добавляет дополнительное измерение к прогностической аналитике IBM SPSS Modeler. Прогностическая аналитика пытается предсказать будущее поведение данных из прошлого, а объектная аналитика направлена на улучшение связности и согласованности текущих данных посредством устранения конфликтов идентичности в самих записях. Идентичность может относиться к индивидууму, организации, а также к любому другому объекту, для которого возможна неоднозначность. Разрешение идентичности может оказаться крайне необходимым для ряда полей, в том числе для управления отношениями с клиентами, обнаружения мошенничества, противодействия отмыванию денег или для национальной и международной безопасности.

IBM SPSS Modeler Social Network Analysis преобразует информацию о взаимосвязях в поля, характеризующие социальное поведение отдельных лиц и групп. При помощи данных, описывающих взаимосвязи, в основе которых лежат социальные сети, IBM SPSS Modeler Social Network Analysis определяет социальных лидеров, влияющих на поведение других участников сети. Кроме того, вы можете определить, какие люди наиболее подвержены влиянию других участников сети. Сочетая полученные результаты с результатами других измерений, можно создать исчерпывающие профили отдельных лиц, на которых будут основаны ваши прогнозные модели. Модели, содержащие эту социальную информацию, выполняются лучше моделей, которые ее не содержат.

IBM SPSS Modeler Text Analytics использует новейшие лингвистические технологии и обработку естественного языка (NLP) для быстрой обработки самых разнообразных неструктурированных текстовых данных, для извлечения и организации ключевых понятий и группирования этих понятий в категории. Извлеченные понятия и категории можно сочетать с существующими структурированными данными, такими как демографические, и применять к моделированию при помощи полного комплекта инструментов исследования данных IBM SPSS Modeler для получения более качественных и специализированных решений.

Документация IBM SPSS Modeler

Документация в формате электронной справки доступна через меню Справка SPSS Modeler. Сюда входит документация по SPSS Modeler, сервер SPSS Modeler и SPSS Modeler Solution Publisher, а также Руководство по прикладным программам и другие сопроводительные материалы.

Полная документация по каждому продукту (включая указания по установке) доступна в формате PDF в подпапках \Documentation каждого продукта DVD. Документы по установке также можно скачать с веб-сайта по адресу <http://www-01.ibm.com/support/docview.wss?uid=swg27038316>.

Кроме того, документация в обоих этих форматах доступна в Информационном центре SPSS Modeler по адресу <http://publib.boulder.ibm.com/infocenter/spssmodl/v16r0m0/>.

Документация SPSS Modeler Professional

В комплект документации SPSS Modeler Professional (включая указания по установке) входят:

- **Руководство пользователя IBM SPSS Modeler.** Общее введение в использование SPSS Modeler, в том числе о создании потоков данных, обработке пропущенных значений, построению выражений CLEM, работе с проектами и отчетами и составлению пакетов потоков для внедрения в IBM SPSS Collaboration and Deployment Services, прогнозирующие прикладные программы или IBM SPSS Modeler Advantage.
- **Узлы источников, обработки и вывода IBM SPSS Modeler.** Описания всех узлов, служащих для чтения, обработки и вывода данных в различных форматах. По существу это все узлы, кроме узлов моделирования.
- **Узлы моделирования IBM SPSS Modeler.** Описания всех узлов, служащих для создания моделей исследования данных. IBM SPSS Modeler предлагает ряд методов моделирования, взятых из таких областей, как обучение машин, искусственный интеллект и статистика.

- **Руководство по алгоритмам IBM SPSS Modeler.** Описание математических основ методов моделирования, используемых в IBM SPSS Modeler. Это руководство доступно только в формате PDF.
- **Руководство по прикладным программам IBM SPSS Modeler.** Примеры в этом руководстве служат кратким специализированным введением к тем или иным методам и технологиям моделирования. Это руководство доступно также в электронном виде в меню Справка. Дополнительную информацию смотрите в разделе “Примеры прикладных программ”.
- **Сценарии и автоматизация IBM SPSS Modeler.** Информация об автоматизации системы путем создания сценариев, включая сценарии свойств, которые могут использоваться для работы с узлами и потоками.
- **Руководство по внедрению IBM SPSS Modeler .** Информация о выполнении IBM SPSS Modeler потоков и сценариев как шагов обработки заданий под управлением IBM SPSS Collaboration and Deployment Services Deployment Manager.
- **Руководство разработчика IBM SPSS Modeler CLEF .** CLEF предоставляет возможности интеграции с программами других производителей, таких как подпрограммы обработки данных или алгоритмы моделирования, как с узлами в IBM SPSS Modeler.
- **Руководство по исследованию данных в базе данных IBM SPSS Modeler.** Информация о том, как использовать мощности вашей базы данных для повышения производительности и расширения диапазона возможностей анализа с привлечением алгоритмов от сторонних производителей.
- **Руководство администратора и руководство по производительности сервер IBM SPSS Modeler .** Информация о том, как сконфигурировать и администрировать сервер IBM SPSS Modeler.
- **Руководство пользователя по консоли администратора IBM SPSS Modeler .** Информация об установке и использовании пользовательского интерфейса консоли для мониторинга и конфигурирования сервер IBM SPSS Modeler. Консоль реализована как подключаемый модуль прикладной программы Диспетчер развертывания .
- **Руководство по CRISP-DM IBM SPSS Modeler.** Пошаговое руководство к использованию методологии CRISP-DM для исследования данных SPSS Modeler.
- **Руководство пользователя IBM SPSS Modeler Batch .** Полное руководство по использованию IBM SPSS Modeler в пакетном режиме, включая подробности выполнения в пакетном режиме и аргументы командной строки. Это руководство доступно только в формате PDF.

Документация SPSS Modeler Premium

В комплект документации SPSS Modeler Premium (включая указания по установке) входят:

- **IBM SPSS Modeler Entity Analytics Руководство пользователя.** Информация об использовании аналитики объектов совместно с SPSS Modeler, в том числе по установке и конфигурированию репозитория, узлам аналитики объектов и задачам управления.
- **IBM SPSS Modeler Social Network Analysis Руководство пользователя.** Руководство по выполнению анализа социальной сети совместно с SPSS Modeler, включая анализ групп и анализ распространения.
- **Руководство пользователя SPSS Modeler Text Analytics .** Информация об использовании аналитики текстов совместно с SPSS Modeler, в том числе по узлам исследования текстов, интерактивной инструментальной среде, шаблонам и другим ресурсам.
- **Руководство пользователя по консоли администратора IBM SPSS Modeler Text Analytics.** Информация об установке и использовании пользовательского интерфейса консоли для мониторинга и конфигурирования сервер IBM SPSS Modeler для использования совместно с SPSS Modeler Text Analytics . Консоль реализована как подключаемый модуль прикладной программы Диспетчер развертывания .

Примеры прикладных программ

Инструменты исследования данных в SPSS Modeler помогают разрешить широкий спектр деловых и организационных проблем, а примеры прикладных программ предоставляют краткие, целевые введения в конкретные методы и способы моделирования. Используемые здесь наборы данных намного меньше огромных складов данных, которыми управляют некоторые исследователи данных, но применяемые понятия и методы должны масштабироваться до реальных прикладных программ.

Обратиться к примерам можно, выбрав **Примеры прикладных программ** в меню Справка в SPSS Modeler. Файлы данных и потоки примеров устанавливаются в папке *Demos* в каталоге установки продукта. Дополнительную информацию смотрите в разделе “Папка demos”.

Примеры моделирования баз данных. Смотрите эти примеры в руководстве *IBM SPSS Modeler In-Database Mining Guide*.

Примеры сценариев. Смотрите эти примеры в руководстве *IBM SPSS Modeler Scripting and Automation Guide*.

Папка demos

Файлы данных и примеры потоков, используемые с примерами прикладных программ, устанавливаются в папке *Demos* в каталоге установки продукта. К этой папке можно также обратиться из группы программ IBM SPSS Modeler в меню Пуск Windows или, щелкнув по *Demos* в списке недавно использовавшихся каталогов в диалоговом окне Открыть файл.

Глава 2. Рекомендации по архитектуре и аппаратным средствам

Архитектура IBM SPSS Modeler

В этом разделе описывается архитектура сервера IBM SPSS Modeler включая программное обеспечение сервера, программное обеспечения клиента и базу данных. Он содержит информацию об особенностях структуры сервера IBM SPSS Modeler, позволяющих достичь оптимальной производительности, и рекомендации по оптимизации производительности путем выбора аппаратного обеспечения надлежащих размеров. В конце приводится раздел по доступу к данным, в котором описано, где сконфигурировать нужные драйверы ODBC.

Описание архитектуры

сервер IBM SPSS Modeler использует трехярусную защищенную архитектуру. Операции программного обеспечения совместно используются клиентом и компьютерами сервера. Установка и использование сервера IBM SPSS Modeler (в сравнении с автономным IBM SPSS Modeler), особенно при работе с большими наборами данных, дает многочисленные преимущества:

- Сервер IBM SPSS Modeler может работать не только в Windows, но и в UNIX, что дает больше гибкости при выборе среды установки. На любой платформе процессам исследования данных можно выделить наиболее большой и быстрый компьютер сервера.
- Сервер IBM SPSS Modeler оптимизирован для максимальной производительности. Когда операции нельзя перенести в базу данных, сервер IBM SPSS Modeler сохраняет промежуточные результаты в виде временных файлов на диске, а не в оперативной памяти. Поскольку для серверов обычно доступно значительное дисковое пространство, сервер IBM SPSS Modeler может выполнять операции сортировки, объединения и агрегирования для очень больших объемов данных.
- Используя архитектуру клиент-сервер, можно централизовать процессы исследования данных в организации. Централизация помогает формализовать роль исследования данных в ваших бизнес-процессах.
- Используя такие инструменты администратора, как IBM SPSS Modeler Administration Console (поставляется с сервером IBM SPSS Modeler) и IBM SPSS Collaboration and Deployment Services (поставляется отдельно), можно отслеживать процессы исследования данных, обеспечивая доступность соответствующих вычислительных ресурсов. С помощью IBM SPSS Collaboration and Deployment Services можно автоматизировать определенные задачи анализа данных, управлять доступом к моделям данных и совместно использовать результаты в рамках вашей организации.

Компоненты распределенной архитектуры IBM SPSS Modeler показаны на диаграмме "Архитектура сервера IBM SPSS Modeler".

- **IBM SPSS Modeler.** На компьютере конечного пользователя устанавливается клиентское программное обеспечение. Оно содержит пользовательский интерфейс и выводит на экран результаты исследования данных. На клиенте установлено полное программное обеспечение IBM SPSS Modeler, однако когда он соединен с сервером IBM SPSS Modeler для распределенного анализа, его механизм выполнения неактивен. IBM SPSS Modeler работает только в операционных системах Windows.
- **сервер IBM SPSS Modeler.** Программное обеспечение сервера, установленное на компьютере сервера, при наличии сетевой связи и с IBM SPSS Modeler (одним или несколькими экземплярами), и с базой данных. Сервер IBM SPSS Modeler работает как служба (в Windows) или как процесс демона (в UNIX), ожидая подключения клиентов. Он обрабатывает выполнение потоков и сценариев, созданных с помощью IBM SPSS Modeler.
- **Сервер баз данных.** Сервер баз данных может быть хранилищем реальных данных (например, Oracle на большом сервере UNIX) или же, для уменьшения влияния на другие операционные системы, рынком данных на локальном сервере или сервере отдела (например, SQL Server в Windows).

Архитектура сервера IBM SPSS Modeler

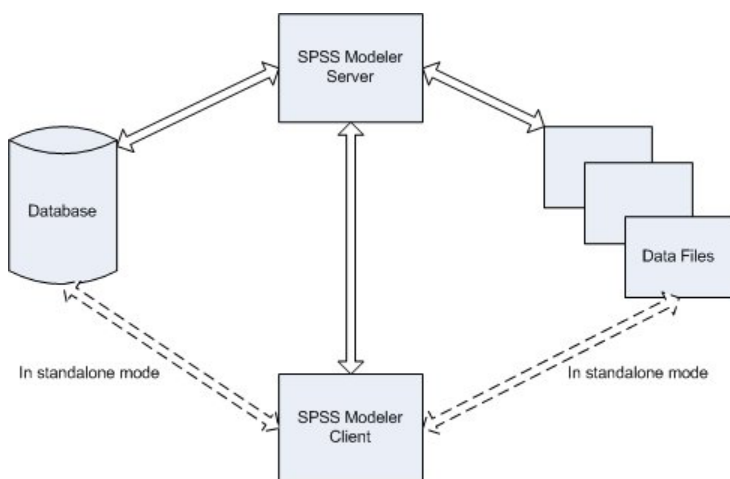


Рисунок 1. Архитектура сервера IBM SPSS Modeler

При распределенной архитектуре большая часть обработки происходит на компьютере сервера. Когда конечный пользователь выполняет поток, IBM SPSS Modeler отправляет на сервер описание этого потока. Сервер определяет, какие операции можно выполнить в SQL, и создает соответствующие запросы. Эти запросы выполняются в базе данных, и полученные в результате данные передаются на сервер для обработки, которую нельзя выполнить с помощью SQL. По завершении обработки обратно на клиент передаются только соответствующие результаты.

При необходимости сервер IBM SPSS Modeler может выполнить все операции IBM SPSS Modeler вне базы данных. Он автоматически балансирует использование оперативной памяти и памяти на дисках для хранения обрабатываемых данных. Этот процесс делает сервер IBM SPSS Modeler полностью совместимым с плоскими файлами.

Балансирование загрузки доступно также путем использования кластера серверов для обработки. Кластеризация доступна, начиная с версии IBM SPSS Collaboration and Deployment Services 3.5, через подключаемый модуль координатора процессов. Дополнительную информацию смотрите в разделе Приложение Е, “Балансировка нагрузок при помощи кластеров серверов”, на стр. 69. С сервером или кластером, управляемым в координаторе процессов, можно соединиться непосредственно при помощи диалогового окна IBM SPSS Modeler Регистрация на сервере. Дополнительную информацию смотрите в разделе “Соединение с сервер IBM SPSS Modeler” на стр. 15.

Автономный клиент

IBM SPSS Modeler можно также сконфигурировать для работы в качестве прикладной программы рабочего стола, показанной ниже на рисунке. Дополнительную информацию смотрите в разделе Глава 3, “Поддержка IBM SPSS Modeler”, на стр. 15.

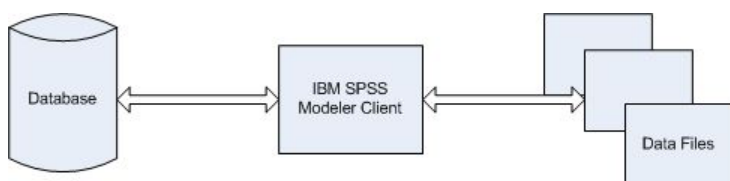


Рисунок 2. Автономный IBM SPSS Modeler

Рекомендации по аппаратным средствам

Планируя установить сервер IBM SPSS Modeler, подумайте, какие аппаратные средства вы будете использовать. Хотя сервер IBM SPSS Modeler предназначен для быстрой работы, его эффективность можно повысить, выбрав оборудование подходящего размера для ваших задач исследования данных. Обновление аппаратных средств - часто наиболее простой и экономичный способ повысить производительность в масштабах организации.

Выделенный сервер. Установите сервер IBM SPSS Modeler на выделенном компьютере сервера, где он не будет конкурировать с другими прикладными программами за ресурсы, включая базы данных, с которыми сервер IBM SPSS Modeler может соединяться. В частности, операции построения моделей ресурсоемки и выполняются намного эффективнее, если не конкурируют при этом с другими прикладными программами.

Примечание: Несмотря на то, можно установить сервер IBM SPSS Modeler на том же компьютере, что и база данных, чтобы уменьшить время передачи данных между базой данных и сервером, помогая избежать сетевых затрат, в большинстве случаев наилучшей конфигурацией является размещение сервера и базы данных на отдельных компьютерах, что устраняет конкуренцию за ресурсы. Для уменьшения стоимости передачи данных между сервером и базой данных необходимо быстрое соединение.

Процессоры. Число процессоров в компьютере должно быть не меньше числа параллельных задач (одновременно выполняемых потоков), которые вы планируете выполнять регулярно. В целом, чем больше процессоров, тем лучше.

- Один экземпляр сервер IBM SPSS Modeler способен принимать требования соединений от нескольких клиентов (пользователей), а каждое клиентское соединение может инициировать выполнение нескольких потоков. Поэтому на одном сервере в любой момент может выполняться несколько задач.
- Как показывает опыт, одного процессора достаточно для одного-двух пользователей, двух процессоров - максимум для четырех пользователей и четырех процессоров - максимум для восьми пользователей. В зависимости от сложности работы, добавьте к этому один дополнительный процессор на каждые двух-четыре пользователей.
- Насколько позволяет возможность передавать часть заданий обработки обратно в базу данных посредством оптимизации SQL, процессор может использоваться совместно двумя или несколькими пользователями с минимальной потерей производительности.
- Возможности многопоточности позволяют при выполнении одной задачи реализовывать преимущества нескольких процессоров, поэтому добавляя процессоры, можно повысить производительность даже при отсутствии параллельных задач. Как правило, многопоточность используется при построении моделей C5.0 и некоторых операций подготовки данных (сортировка, агрегирование и объединение).

64-битные платформы. Если вы планируете обрабатывать или строить модели на очень больших объемах данных, используйте в качестве платформы сервера IBM SPSS Modeler 64-битную Solaris, Windows или Linux и максимизируйте объем оперативной памяти для компьютера. Для более крупных наборов данных сервер может быстро исчерпать пределы выделяемой на процесс памяти, накладываемые 32-битными платформами, вынуждая сбрасывать данные на диск и тем самым значительно увеличивая время выполнения. В 64-битных реализациях сервера полезна дополнительная оперативная память, рекомендуется как минимум 8 Гбайт. 64-битная поддержка доступна для платформ Solaris, Windows и Linux.

Будущие потребности. По возможности, всегда старайтесь проверять, что оборудование сервера было расширяемо в смысле добавления памяти и процессоров, чтобы соответствовать требованиям возросшей нагрузки (например, при увеличении числа параллельных пользователей или увеличении потребности в обработке у существующих пользователей), а также развития возможностей многопоточности на сервере IBM SPSS Modeler в будущем.

Временное дисковое пространство и требования к оперативной памяти

Для обработки больших объемов данных сервер IBM SPSS Modeler использует временное дисковое пространство. Требуемый объем временного пространства зависит от объема обрабатываемых данных и типа выполняемых операций. Объем данных пропорционален количеству строк и количеству столбцов. Чем больше строк и столбцов вы обрабатываете, тем больше дискового пространства потребуется.

В этом разделе описываются условия, при которых требуются временное дисковое пространство и дополнительная оперативная память, и объясняется, как оценить их потребный объем. Обратите внимание на то, что в этом разделе не обсуждаются требования к временному дисковому пространству для процессов, выполняемых в базе данных, так как эти требования специфичны для каждой базы данных.

Ситуации, требующие временного дискового пространства

На сервер IBM SPSS Modeler есть мощная функция оптимизации SQL, которая позволяет, когда это возможно, выполнять обработку в базе данных, а не на сервере. Однако оптимизацию SQL нельзя использовать в любого из следующих ситуаций:

- Обрабатываемые данные находятся в плоском файле, а не в базе данных.
- Оптимизация SQL отключена.
- Операцию обработки нельзя оптимизировать при помощи SQL.

Если оптимизацию SQL нельзя использовать, перечисленные ниже узлы преобразования данных и функции CLEM создадут временные дисковые копии всех или части данных. Если потоки, используемые на вашем сайте, содержат эти команды или функции обработки, вам, возможно, придется выделить дополнительное дисковое пространство на вашем сервере.

- Узел агрегации
- Особый узел
- узел разделения на интервалы
- Узел слияния при использовании опции слияния по ключу
- Любой узел моделирования
- Узел сортировки
- Узел табличного вывода
- Функции @OFFSET, в которых условие поиска использует @THIS.
- Любая функция @, например, @MIN, @MAX или @AVE, в которой вычисляется параметр смещения.

Вычисление объема временного дискового пространства

В целом сервер IBM SPSS Modeler должен иметь возможность записать временный файл, который по размеру, как минимум, *в три раза превышает* исходный набор данных. Например, если размер файла данных - 2 Гбайта, и генерирование SQL не используется, сервер IBM SPSS Modeler потребуется для обработки данных 6 Гбайт дискового пространства. Поскольку для каждой параллельной учетной записи пользователя создаются свои собственные временные файлы, необходимо соответственно увеличить дисковое пространство для всех одновременно работающих пользователей.

Если оказывается, что на вашем сайте часто используются большие временные файлы, попробуйте использовать отдельную файловую систему для временных файлов IBM SPSS Modeler, созданных на отдельном диске. Наилучшие результаты дает использование RAID 0 или многополосного набора данных, охватывающего несколько физических дисков, для ускорения дисковых операций, лучше всего с отдельным контроллером для каждого диска в многополосной файловой системе.

Требования к оперативной памяти

Для большей части задач обработки, которые нельзя выполнить в базе данных, сервер IBM SPSS Modeler сохраняет промежуточные результаты как временные файлы на диске, а не в оперативной памяти. Однако при моделировании узлов память по возможности используется. Для узлов нейросети, Коонена и k-средних

требуются большие объемы оперативной памяти. Если эти узлы часто используются на вашем сайте, имеет смысл установить на сервере больше оперативной памяти.

В целом необходимый объем оперативной памяти оценивается по формуле
(число_записей * число_ячеек_на_запись) * число_байтов_на_ячейку

где число_ячеек_на_запись может быть очень большим при наличии номинальных полей.

Текущие рекомендации по оперативной памяти смотрите в разделе системных требований руководства по установке сервера. Для четырех или более одновременно работающих пользователей рекомендуется еще больше памяти. Память должна совместно использоваться параллельными задачами, поэтому ее объем нужно соответственно увеличить. В целом добавление памяти, вероятно, будет одним из самых эффективных по затратам способов улучшить производительность по организации.

Доступ к данным

Для чтения или записи данных из базы данных пользователь должен установить источник данных ODBC, настроить соответствующую базу данных и установить разрешения на чтение и запись. IBM SPSS Data Access Pack содержит набор драйверов ODBC, который может использоваться для этой цели; эти драйверы доступны на IBM SPSS Data Access Pack DVD или на сайте скачивания. Для получения дополнительных сведений о создании и настройке разрешений для источников данных ODBC обратитесь к своему администратору базы данных.

Поддерживаемые драйверы ODBC

Наиболее свежие сведения о базах данных и драйверах ODBC, работоспособность которых проверена и поддерживается при использовании с IBM SPSS Modeler 16, находятся в матрице совместимости программного обеспечения на веб-сайте технической поддержки (<http://www.ibm.com/support>).

Установка драйверов на компьютеры

Следует заметить, что драйверы ODBC должны быть установлены и настроены на каждом из компьютеров, обрабатывающих данные.

- Если программа IBM SPSS Modeler используется в локальном (отдельном) режиме, драйверы должны быть установлены на локальный компьютер.
- Если программы IBM SPSS Modeler работают в распределенном режиме совместно с сервер IBM SPSS Modeler, драйверы ODBC должны быть установлены на компьютере с сервер IBM SPSS Modeler. Для сервер IBM SPSS Modeler в системах UNIX смотрите также "Конфигурирование драйверов ODBC в системах UNIX" далее в этом разделе.
- При необходимости доступа к одним и тем же источникам данных как с IBM SPSS Modeler, так и с сервер IBM SPSS Modeler, драйверы ODBC должны быть установлены на обоих компьютерах.
- При использовании IBM SPSS Modeler совместно со службами терминалов драйверы ODBC должны быть установлены на сервере служб терминалов, на котором установлена программа IBM SPSS Modeler.

Важно: Если для доступа к базе данных Teradata в UNIX используется сервер IBM SPSS Modeler и вы используете драйвер ODBC, поставляемый с Teradata, необходимо использовать менеджер драйверов ODBC, который устанавливается вместе с этим драйвером. (**Примечание:** Не нужно вносить эти изменения, если вы используете драйвер SDAP Teradata). Чтобы изменить на сервер IBM SPSS Modeler, укажите значение ODBC_DRIVER_MANAGER_PATH рядом с верхней частью сценария modelersrv.sh, оформив его в виде комментариев. Данная переменная среды должна указывать на расположение менеджера драйверов ODBC, который поставляется вместе с драйвером ODBC базы данных Teradata (/usr/odbc/lib в установке по умолчанию драйвера ODBC базы данных Teradata). Необходимо перезапустить сервер IBM SPSS Modeler, чтобы изменения вступили в силу. Для получения дополнительной информации по платформам сервер IBM SPSS Modeler, которые предлагают поддержку для доступа к базе данных и поддерживаемую версию драйвера ODBC базы данных, зайдите на корпоративный сайт поддержки на <http://www.ibm.com/support>.

Настройка драйверов ODBC в системах UNIX

По умолчанию менеджер драйверов DataDirect не настроен для систем сервер IBM SPSS Modeler UNIX. Чтобы настроить UNIX для загрузки менеджера драйверов DataDirect, введите следующие команды:

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

В результате будет удалена ссылка по умолчанию и создана ссылка на диспетчер устройств DataDirect.

Примечание: Вышеупомянутые правила применяются специально к доступу к данным в базе данных. Другие типы операций с файлами, например, открытие и сохранение файлов потоков, проектов, моделей, узлов, PMML, вывода и сценария, всегда выполняются на клиенте и всегда определяются в терминах файловой системы клиентского компьютера. Кроме того, команда IBM SPSS Modeler Задать каталог задает рабочий каталог для *локальных* клиентских объектов (например, потоков), но не влияет на рабочий каталог сервера.

UNIX. Информацию о том, как сконфигурировать сервер IBM SPSS Modeler в UNIX для работы с технологией доступа к данным IBM SPSS Statistics, смотрите в разделе Приложение В, “Конфигурирование сценариев запуска UNIX”, на стр. 55.

Ссылки на файлы данных

Windows. Если вы храните данные на том же компьютере, что сервер IBM SPSS Modeler, рекомендуется указать путь к данным относительно компьютера сервера (например, *C:\ServerData\Sales 1998.csv*). Когда для поиска файла не используется сеть, производительность выше.

Если данные хранятся на другом хосте, рекомендуется использовать ссылки на файлы в формате UNC (например, *\\mydataserver\ServerData\Sales 1998.csv*). Обратите внимание на то, что имена UNC работают, только если путь содержит имя совместно используемого сетевого ресурса. У компьютера, с которого выполняется обращение, должно быть разрешение на чтение указанного файла. Если вы часто переключаетесь с распределенного режима на режим локального анализа, используйте ссылки на файлы UNC, поскольку они работают независимо от режима.

UNIX. Для ссылки на файлы данных на сервере UNIX задайте полную спецификацию файлов, используя прямую дробную черту (например, */public/data/ServerData/Sales 1998.csv*). Избегайте символа обратной дробной черты в каталоге UNIX и в именах файлов данных, используемых с сервером IBM SPSS Modeler. Не имеет значения, использует ли текстовый файл формат UNIX или DOS - оба формата обрабатываются автоматически.

Импорт файлов данных IBM SPSS Statistics

Если на вашем сайте запущен также сервер IBM SPSS Statistics пользователям может понадобиться импортировать или экспортировать данные IBM SPSS Statistics, работая в распределенном режиме. Вспомним, что при работе в распределенном режиме IBM SPSS Modeler представляет файловую систему сервера. Клиент IBM SPSS Statistics работает аналогичным образом. Чтобы между двумя программами был возможен импорт и экспорт, оба клиента должны работать в одинаковом режиме. В противном случае их представления файловых систем будут различаться, и они не будут в состоянии совместно использовать файлы. Узлы IBM SPSS Statistics IBM SPSS Modeler могут автоматически запустить клиент IBM SPSS Statistics, но пользователям вначале необходимо убедиться, что клиент IBM SPSS Statistics работает в том же режиме, что и IBM SPSS Modeler.

Инструкции по установке

Информацию о том, как установить сервер IBM SPSS Modeler, смотрите в инструкциях в папке *\documentation\installation* на компакт-диске сервера. Для Windows и UNIX предназначены отдельные документы.

Полную информацию по установке и использованию IBM SPSS Modeler смотрите на компакт-диске клиента.

Глава 3. Поддержка IBM SPSS Modeler

Этот раздел предназначен для администраторов и персонала справочной службы, поддерживающих пользователей IBM SPSS Modeler. Он охватывает следующие темы:

- Как войти в систему сервер IBM SPSS Modeler (или работать автономно, отсоединившись от сервера)
- Данные и файловые системы, которые могут понадобиться пользователям
- Учетные записи пользователей и режим доступа к файлам сервера IBM SPSS Modeler
- Различия результатов, которые могут видеть пользователи, переключаясь между сервером IBM SPSS Modeler и IBM SPSS Modeler

Соединение с сервер IBM SPSS Modeler

IBM SPSS Modeler можно запускать как автономную прикладную программу или как клиент, подключенный непосредственно к сервер IBM SPSS Modeler или к сервер IBM SPSS Modeler или кластеру сервера через подключаемый модуль координатора процессов в IBM SPSS Collaboration and Deployment Services. В нижней части окна IBM SPSS Modeler слева будет выведено текущее состояние соединения.

Каждый раз, когда требуется соединиться с сервером, можно вручную ввести имя сервера, к которому вы хотите подключиться, или выбрать заранее заданное вами имя. Однако если используется IBM SPSS Collaboration and Deployment Services, можно выполнить поиск в списке серверов или кластеров сервера в диалоговом окне Регистрация на сервере. Возможность просмотра через службы Statistics, работающие в сети, обеспечивается координатором процессов.

Чтобы соединиться с сервером:

1. В меню Инструменты выберите **Регистрация на сервере**. Откроется диалоговое окно Регистрация на сервере. Другой вариант - дважды щелкните по области состояния соединения в окне IBM SPSS Modeler.
2. В диалоговом окне задайте опции соединения с компьютером локального сервера или выберите соединение в таблице.
 - Нажмите кнопку **Добавить** или **Изменить**, чтобы добавить или изменить соединение. Дополнительную информацию смотрите в разделе “Добавление и изменение соединений с сервер IBM SPSS Modeler” на стр. 17.
 - Нажмите кнопку **Поиск** для доступа к серверу или кластеру сервера в координаторе процессов. Дополнительную информацию смотрите в разделе “Поиск серверов в IBM SPSS Collaboration and Deployment Services” на стр. 18.

Таблица серверов. Эта таблица содержит набор заданных соединений с серверами. В таблице выводятся соединение по умолчанию, имя сервера, описание и номер порта. Можно вручную добавить новое соединение, а также выбрать или найти существующее соединение. Чтобы задать конкретный сервер в качестве соединения по умолчанию, включите переключатель в столбце По умолчанию таблицы этого соединения.

Путь к данным по умолчанию. Задайте путь для данных на компьютере сервера. Нажмите кнопку с многоточием (...) для просмотра нужного положения.

Задать параметры авторизации. Оставьте этот переключатель выключенным, чтобы разрешить возможность **единой регистрации**, которая пытается зарегистрироваться на сервере, используя локальное имя пользователя и пароль на вашем компьютере. Если единая регистрация невозможна или отключена этим переключателем (например, для входа в учетную запись администратора), для ввода параметров регистрации будут доступны следующие поля.

ID пользователя. Введите имя пользователя для входа на сервер.

Пароль. Укажите пароль, связанный с указанным имени пользователя.

Домен. Задайте домен, который будет использоваться для входа на сервер. Доменное имя требуется, только если компьютер сервера находится в другом домене Windows по отношению к компьютеру клиента.

3. Нажмите кнопку **ОК**, чтобы завершить подключение.

Чтобы отсоединиться от сервера:

1. В меню Инструменты выберите **Регистрация на сервере**. Откроется диалоговое окно Регистрация на сервере. Другой вариант - дважды щелкните по области состояния соединения в окне IBM SPSS Modeler.
2. В диалоговом окне выберите Локальный сервер и нажмите кнопку **ОК**.

Конфигурирование единой регистрации

Вы можете соединиться с сервером IBM SPSS Modeler, работающим на любой поддерживаемой платформе, при помощи единой регистрации. Для соединения с помощью единой регистрации надо сначала сконфигурировать сервер IBM SPSS Modeler и клиентские компьютеры.

Если вы используете единую регистрацию для соединения с сервером IBM SPSS Modeler и IBM SPSS Collaboration and Deployment Services, надо соединиться с IBM SPSS Collaboration and Deployment Services до соединения с IBM SPSS Modeler.

Примечание: Прежде чем конфигурировать компьютеры сервера и клиентов IBM SPSS Modeler для единой регистрации, надо убедиться, что у этих компьютеров есть доступ к серверу контроллера домена.

Конфигурирование единой регистрации в Windows

Чтобы сконфигурировать сервер IBM SPSS Modeler

1. Убедитесь, что компьютер сервера Windows входит в домен Active Directory (AD).
2. В положении установки сервер IBM SPSS Modeler найдите папку с именем config.
3. Создайте в папке config подпапку с именем sso.
4. В папке sso создайте файл krb5.conf. Инструкции по созданию файла krb5.conf смотрите по адресу http://web.mit.edu/kerberos/krb5-current/doc/admin/conf_files/krb5_conf.html. Пример файла krb5.conf приводится ниже:

```
[libdefaults]
    default_realm = MODELERSSO.COM
    dns_lookup_kdc = true
    dns_lookup_realm = false

[realms]
    MODELERSSO.COM = {
        kdc = <AD server IP>:88
        admin_server = <AD server IP>:749
        default_domain = MODELERSSO.COM
    }

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

Чтобы сконфигурировать клиент IBM SPSS Modeler

1. Убедитесь, что локальный компьютер Windows, на котором работает IBM SPSS Modeler, входит в домен AD.
2. Добавьте пользователя домена в качестве администратора локального компьютера.
3. Разрешите Windows обращаться к ключу сеанса TGT:
 - a. В меню **Пуск** выберите **Выполнить**.
 - b. Введите regedit и нажмите кнопку **ОК**, чтобы открыть Редактор реестра.
 - c. Найдите положение реестра, соответствующее операционной системе локального компьютера:

- В Windows XP: My Computer\HKEY_LOCAL_MACHINE\System\CurrentControlSet\Control\Lsa\Kerberos
 - В Windows Vista или Windows 7: My Computer\HKEY_LOCAL_MACHINE\System\CurrentControlSet\Control\Lsa\Kerberos\Parameters
- d. Щелкните правой кнопкой мыши по папке и выберите **Создать > Параметр DWORD**. Новое значение надо назвать allowtgtsessionkey.
 - e. Задайте для allowtgtsessionkey шестнадцатеричное значение 1, то есть 0x0000001.
 - f. Закройте Редактор реестра.
 - g. Если операционная система на локальном компьютере - Windows Vista или Windows 7, надо запустить kinit.exe, которую можно найти в каталоге <IBM SPSS Modeler каталог установки>\jre\bin.
4. В папке config положения установки IBM SPSS Modeler создать папку с именем sso.
 5. Скопировать файл krb5.conf с сервера в папку sso.
 6. Перезапустить локальный компьютер и компьютер сервера.

Конфигурирование единой регистрации в UNIX

Чтобы сконфигурировать единую регистрацию на компьютере сервера UNIX, можно добавить компьютер UNIX в домен AD Windows, а затем следовать инструкциям по конфигурированию единой регистрации в Windows. Дополнительную информацию смотрите в разделе “Конфигурирование единой регистрации в Windows” на стр. 16. Другой вариант - выполнить следующие действия:

1. Создать учетную запись пользователя пользователя домена для компьютера UNIX.
2. Изменить имя хоста. Если вы используете RedHat Linux, откройте файл /etc/sysconfig/network и измените HOSTNAME, задав его в форме <имя>.<царство>. Это позволит AD найти регистрационные данные сервера.
3. Чтобы позволить серверу DNS найти компьютер UNIX, выполните одно из следующих действий:
 - Откройте файл %windows%/system32/drivers/etc/hosts и добавьте в него отображение IP/хост, например:
192.168.1.102 test.modelersso.com test
либо
 - Добавьте новую запись зоны обратного просмотра. Это добавит отображение IP/хост на сервер DNS.

Если запись DNS для компьютера UNIX задана неправильно, вы можете вручную добавить запись обратного просмотра на сервер DNS.
4. Следуйте инструкциям по конфигурированию клиента IBM SPSS Modeler в теме “Конфигурирование единой регистрации в Windows” на стр. 16.

Добавление и изменение соединений с сервер IBM SPSS Modeler

Можно вручную изменить или добавить соединение с сервером в диалоговом окне Регистрация на сервере. При нажатии кнопки Добавить открывается пустое диалоговое окно Добавить/Изменить сервер, в которое можно ввести подробности соединения с сервером. После выбора существующего соединения и нажатия кнопки Изменить в диалоговом окне Подключение к серверу откроется диалоговое окно Добавить/Изменить сервер со сведениями об этом соединении, где можно внести необходимые изменения.

Примечание: Соединение с сервером, добавленное из IBM SPSS Collaboration and Deployment Services, нельзя изменить, поскольку имя, номер порта и другая информация заданы в IBM SPSS Collaboration and Deployment Services.

Чтобы добавить соединения с сервером

1. В меню Инструменты выберите **Регистрация на сервере**. Откроется диалоговое окно Регистрация на сервере.
2. В этом диалоговом окне нажмите кнопку **Добавить**. Откроется диалоговое окно Добавить/Изменить сервер.

3. Введите сведения о соединении с сервером и нажмите кнопку **ОК**, чтобы сохранить это соединение и вернуться в диалоговое окно Регистрация на сервере.
- **Сервер.** Задайте доступный сервер или выберите его из списка. Компьютер сервера задается алфавитно-цифровым именем (например, *myserver*) или IP-адресом, назначенным компьютеру сервера, (например, 202.123.456.78).
 - **Порт.** Укажите номер порта, на котором сервер ожидает сообщений. Если номер по умолчанию не работает, попросите у администратора системы правильный номер порта.
 - **Описание.** Введите необязательное описание для этого соединения с сервером.
 - **Обеспечить защищенное соединение (использовать SSL).** Укажите, нужно ли использовать подключение по протоколу SSL (**Secure Sockets Layer**). SSL - обычно используемый протокол для защиты данных, пересылаемых по сети. Для доступа к этой функции необходимо разрешить SSL на сервере, где запущен сервер IBM SPSS Modeler. При необходимости уточните подробности у вашего администратора.

Чтобы изменить соединения с сервером

1. В меню Инструменты выберите **Регистрация на сервере**. Откроется диалоговое окно Регистрация на сервере.
2. В этом диалоговом окне выберите соединение, которое нужно изменить, и нажмите кнопку **Изменить**. Откроется диалоговое окно Добавить/Изменить сервер.
3. Измените сведения о соединении с сервером и нажмите кнопку **ОК**, чтобы сохранить это соединение и вернуться в диалоговое окно Регистрация на сервере.

Поиск серверов в IBM SPSS Collaboration and Deployment Services

Вместо того, чтобы вводить соединение с сервером вручную, можно выбрать в сети сервер или кластер серверов через координатор процессов, доступный в IBM SPSS Collaboration and Deployment Services. Кластер серверов - это группа серверов, внутри которой координатор процессов определяет сервер, лучше всего подходящий для ответов на запрос обработки.

Хотя можно и вручную добавлять серверы в диалоговом окне Подключение к серверу, поиск имеющихся серверов позволяет соединиться с серверами без необходимости знания правильного имени сервера и номера порта. Эта информация предоставляется автоматически. Однако при этом необходима правильная информация об учетной записи, такая как имя пользователя, домен и пароль.

Примечание: Если у вас нет доступа к функции координатора процессов, можно вручную ввести имя сервера с которым вы хотите соединиться, или выбрать имя, заданное вами ранее. Дополнительную информацию смотрите в разделе “Добавление и изменение соединений с сервер IBM SPSS Modeler” на стр. 17.

Чтобы найти серверы и кластеры:

1. В меню Инструменты выберите **Регистрация на сервере**. Откроется диалоговое окно Регистрация на сервере.
2. В диалоговом окне нажмите кнопку **Поиск**, чтобы открыть диалоговое окно Поиск серверов. Если вы еще не вошли в IBM SPSS Collaboration and Deployment Services при попытке просмотра координатора процессов, вам предложат сделать это.
3. Выберите сервер или кластер серверов в списке.
4. Нажмите кнопку **ОК**, чтобы закрыть диалоговое окно и добавить это соединение в диалоговое окно Подключение к серверу.

Данные и файловые системы

Пользователям, работающим с сервером IBM SPSS Modeler, вероятно, понадобится обращаться к файлам данных и другим сетевым источникам данных, а также сохранять файлы в сети. В зависимости от обстоятельств, им может потребоваться следующая информация:

- **Информация об источнике данных ODBC.** Если пользователям нужен доступ к источникам данных ODBC, определенным на компьютере сервера, им понадобятся имена, описания и информация регистрации (включая ID и пароли для входа в систему базы данных) для источников данных.
- **Доступ к файлам данных.** Если пользователям нужен доступ к файлам данных на компьютере сервера или в другом месте сети, им понадобятся имена и сведения о положении файлов данных.
- **Положение сохраненных файлов.** Если пользователи сохраняют данные, будучи подключенными к серверу IBM SPSS Modeler, они могут попытаться сохранить файлы на компьютере сервера. Однако запись в это положение часто бывает не разрешена. В этом случае сообщите пользователям, где следует сохранять файлы с данными. (Как правило, это домашний каталог пользователя.)

Аутентификация пользователей

Для аутентификации пользователей, соединяющихся с сервером, сервер IBM SPSS Modeler использует средства оперативной системы компьютера сервера. Когда пользователь соединяется с сервером IBM SPSS Modeler, все операции, выполняемые от его имени, выполняются в контексте защиты этого пользователя. Доступ к таблицам базы данных осуществляется с использованием привилегий для пользователей и/или паролей в самой базе данных.

Windows. В Windows войти в систему может любой пользователь с действительной учетной записью в сети хоста. При аутентификации по умолчанию необходимо изменить права доступа пользователей к каталогу `<каталог_установки_сервера_modeler>\Tmp`. Без этих прав пользователи не смогут войти на сервер IBM SPSS Modeler с клиента, используя аутентификацию по умолчанию в Windows.

UNIX. По умолчанию предполагается, что сервер IBM SPSS Modeler работает в UNIX как root. Это позволяет входить в систему любому пользователю с действительной учетной записью в сети хоста и ограничивает доступ пользователей к их собственным файлам и каталогам. Однако можно сконфигурировать сервер IBM SPSS Modeler для работы без привилегии root. В этом случае необходимо создать базу данных секретных паролей, которая будет использоваться для аутентификации, а все пользователи IBM SPSS Modeler должны будут совместно использовать единственную учетную запись пользователя UNIX (и, следовательно, будут иметь совместный доступ к файлам данных). Дополнительную информацию смотрите в разделе “Конфигурирование в качестве рядового пользователя (не root) с помощью базы данных секретных паролей” на стр. 64.

На платформах Solaris, HP-UX, Linux и AIX сервер IBM SPSS Modeler использует для аутентификации PAM. Если требуется, используйте имя службы `modelerserver`, чтобы сконфигурировать модули PAM для сервера IBM SPSS Modeler.

Создание файлов

Когда сервер IBM SPSS Modeler обращается к данным для их обработки, ему часто приходится создавать на диске временную копию этих данных. Объем дискового пространства, выделяемого для временных файлов, зависит от размера файла с данными, с которым работает конечный пользователь, и от типа выполняемого анализа. Дополнительную информацию смотрите в разделе “Временное дисковое пространство и требования к оперативной памяти” на стр. 10.

UNIX. В версиях сервера IBM SPSS Modeler UNIX для задания режима доступа к временным файлам используется команда UNIX `umask`. Вы можете переопределить разрешения для сервера по умолчанию. Дополнительную информацию смотрите в разделе “Управление разрешениями при создании файла” на стр. 56.

Различия в результатах

Пользователи, выполняющие анализ в обоих режимах, могут увидеть незначительные различия в результатах между IBM SPSS Modeler и сервером IBM SPSS Modeler. Такое несоответствие обычно объясняется различиями в порядке записей или режиме округления.

Порядок записей. Если поток явно не упорядочивает записи путем их сортировки, порядок, в котором представлены записи, может различаться для потоков, выполняемых локально и выполняемых на сервере. Возможны также различия в порядке между операциями, выполняемыми в базе данных и на сервере IBM SPSS Modeler. Эти различия вызваны различиями алгоритмов, используемых каждой системой для реализации таких функций, как агрегация, которые могут переупорядочивать записи. Кроме того, обратите внимание на то, что SQL не определяет порядок, в котором записи возвращаются из базы данных, в тех случаях, когда операция упорядочивания не была выполнена явно.

Различия в округлении. IBM SPSS Modeler в локальном режиме использует другой внутренний формат для хранения значений с плавающей точкой, чем сервер IBM SPSS Modeler. Из-за различий в округлении обе версии могут давать несколько разные результаты.

Глава 4. Администрирование сервера IBM SPSS Modeler

Эта глава содержит информацию о запуске и остановке сервера IBM SPSS Modeler конфигурировании различных опций сервера и интерпретации файла журнала. В ней описано, как использовать IBM SPSS Modeler Administration Console, прикладную программу, помогающую конфигурировать сервер и отслеживать его работу. Инструкции по установке для этого компонента смотрите в инструкциях по установке сервера IBM SPSS Modeler, доступных с этим продуктом.

Запуск и остановка сервера IBM SPSS Modeler

Сервер IBM SPSS Modeler работает как служба в Windows и как процесс демона в UNIX.

Замечание по планированию: Остановка сервера IBM SPSS Modeler приводит к отключению конечных пользователей и завершению их сеансов, поэтому желательно планировать перезапуски сервера на периоды низкой нагрузки. Если это невозможно, не забудьте уведомить пользователей прежде, чем остановить сервер.

Для запуска, остановки и проверки состояния в Windows

В Windows управление сервером IBM SPSS Modeler осуществляется через диалоговое окно Службы на Панели управления Windows.

1. **Windows XP.** Откройте меню Пуск Windows. Выберите **Панель управления**. Дважды щелкните по значку **Администрирование** и затем по значку **Службы**.
Windows 2003 или 2008. Откройте меню Пуск Windows. Выберите **Панель управления**, затем **Администрирование**, затем **Службы**.
2. Выберите службу **Сервер IBM SPSS Modeler <nn.n>**. Теперь можно проверить его состояние, запустить или остановить его, а также изменить параметры запуска, как вам требуется.

По умолчанию служба сконфигурирована для автоматического запуска, что означает, что в случае остановки она автоматически перезапустится при перезагрузке компьютера. При таком запуске служба работает в автономном режиме, и можно выйти из системы на компьютере сервера, не влияя на нее.

Для запуска, остановки и проверки состояния в UNIX

В UNIX вы запускаете или останавливаете сервер IBM SPSS Modeler, запуская сценарий *modelersrv.sh* в каталоге установки этого сервера.

1. Перейдите в каталог, где установлен сервер IBM SPSS Modeler. Например, введите в командной строке UNIX
`cd /usr/modelersrv`
где *modelersrv* - каталог, где установлен сервер IBM SPSS Modeler.
2. Чтобы запустить сервер, введите в командной строке
`./modelersrv.sh start`
3. Чтобы остановить сервер, введите в командной строке
`./modelersrv.sh stop`
4. Чтобы проверить состояние сервера IBM SPSS Modeler введите в командной строке UNIX
`./modelersrv.sh list`
и посмотрите вывод (он аналогичен выводу команды UNIX `ps`). Первый процесс в списке - это процесс демона сервера IBM SPSS Modeler а остальные процессы - это сеансы IBM SPSS Modeler.

Программа установки сервера IBM SPSS Modeler содержит сценарий (*auto.sh*), который конфигурирует систему, чтобы демон сервера запускался автоматически во время загрузки. Если вы запустили этот сценарий и затем остановили сервер, демон сервера автоматически перезапустится при перезагрузке компьютера. Дополнительную информацию смотрите в разделе “Автоматический запуск и остановка сервера IBM SPSS Modeler” на стр. 55.

Ограничения ядра UNIX

Вы должны убедиться в том, что ограничения ядра в системе допускают работу сервера IBM SPSS Modeler. В частности, важно отсутствие ограничений для данных, памяти и файлов, и они должны быть сняты в среде сервера IBM SPSS Modeler. Для этого:

1. Добавьте следующие команды в *modelersrv.sh*:

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

2. Перезапустите сервер IBM SPSS Modeler.

Обработка неотвечающих процессов сервера (системы UNIX)

Процессы сервера IBM SPSS Modeler могут перестать отвечать по нескольким причинам, включая ситуации, когда они выполняют системный вызов или вызов драйвера ODBC, который блокируется (вызов не возвращает управление или возвращает его спустя очень длительное время). Когда процессы UNIX оказываются в этом состоянии, их можно убрать с помощью команды UNIX *kill* (прерывания, инициированные клиентом IBM SPSS Modeler, или закрытие клиента IBM SPSS Modeler не принесут результата). Команда *kill* - альтернатива обычной команде *stop*; администратор может использовать для ввода команды *kill* файл *modelersrv.sh*.

В системах, чувствительных к накоплению неиспользуемых процессов сервера ("зомби-процессов"), рекомендуется останавливать и перезапускать сервер IBM SPSS Modeler через правильные промежутки времени, используя последовательность команд:

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

Процессы IBM SPSS Modeler, которые были завершены с помощью команды *modelersrv.sh kill*, оставляют после себя временные файлы (во временном каталоге), которые нужно будет удалить вручную. Временные файлы могут остаться также в некоторых других ситуациях, например, при аварийном завершении прикладных программ из-за недостатка ресурсов, вмешательства пользователей, отказов системы или других причин. Поэтому, если вы периодически перезапускаете сервер IBM SPSS Modeler, рекомендуется удалять из временного каталога IBM SPSS Modeler все образовавшиеся там файлы.

После закрытия всех процессов сервера и удаления временных файлов сервера IBM SPSS Modeler можно безопасно перезапустить.

Администрирование

У сервера IBM SPSS Modeler есть несколько конфигурируемых опций, управляющих его поведением. Эти опции можно задать двумя способами:

- Воспользуйтесь прикладной программой IBM SPSS Modeler Administration Console, которая доступна бесплатно покупателям IBM SPSS Modeler. Дополнительную информацию смотрите в разделе “Использование IBM SPSS Modeler Administration Console” на стр. 23.
- Используйте текстовый файл *options.cfg* в каталоге *[путь_установки_сервера]/config*. Дополнительную информацию смотрите в разделе “Использование файла *options.cfg*” на стр. 29.

Рекомендуется не редактировать файл *options.cfg*, а установить IBM SPSS Modeler Administration Console и использовать его в качестве инструмента администрирования. Редактирование файла требует доступа к файловой системе сервера IBM SPSS Modeler, но IBM SPSS Modeler Administration Console позволяет авторизовать любого пользователя для настройки этих опций. Кроме того, IBM SPSS Modeler Administration Console содержит дополнительную информацию о процессах сервера, позволяя отслеживать его использование и производительность. В отличие от редактирования файла конфигурации, большинство опций конфигурации можно изменить, не перезапуская сервер IBM SPSS Modeler.

Дополнительную информацию об использовании IBM SPSS Modeler Administration Console и файла *options.cfg* смотрите в следующих разделах.

Использование IBM SPSS Modeler Administration Console

Программа Modeler Administration Console предоставляет пользовательский интерфейс консоли для мониторинга и конфигурирования ваших установок серверов SPSS Modeler; она доступна без оплаты для текущих заказчиков серверов SPSS Modeler. Эту прикладную программу можно установить только на компьютерах Windows; однако она может управлять сервером на любой поддерживаемой платформе.

Многие опции, доступные через Modeler Administration Console, можно также задать в файле *options.cfg*, расположенном в подкаталоге */config* каталога установки сервера SPSS Modeler. Однако программа Modeler Administration Console предоставляет общий графический интерфейс, который позволяет соединиться со многими серверами, конфигурировать и отслеживать их.

Запуск Modeler Administration Console

В меню Пуск Windows выберите **[Все] Программы**, затем **IBM SPSS Collaboration and Deployment Services**, затем **Deployment Manager** и затем **IBM SPSS Modeler Administration Console <nn.n>**.

При первом запуске программы вы увидите пустые панели Администратор серверов и Свойства (если у вас еще не установлен Deployment Manager с сконфигурированным соединением с IBM SPSS Collaboration and Deployment Services). После конфигурирования Modeler Administration Console на панели Администратор серверов слева выводится узел для каждого сервера SPSS Modeler, который вы хотите администрировать. На панели справа выводятся опции конфигурации для выбранного сервера. Сначала надо задать соединение для каждого сервера, который вы хотите администрировать.

Конфигурирование доступа с Modeler Administration Console

Доступом администратора к серверу SPSS Modeler через Modeler Administration Console управляет строка *administrators* в файле *options.cfg*, расположенном в подкаталоге */config* каталога установки сервера SPSS Modeler. Эта строка по умолчанию закомментирована, так что вам надо отредактировать ее, чтобы разрешить доступ определенным людям, или, задав *, всем пользователям, как показано в следующих примерах:

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

- Эта строка должна начинаться с *administrators,*, а записи надо заключить в кавычки. Эти записи регистрозависимы.
- Отделяйте ID пользователей друг от друга запятыми.
- Для учетных записей Windows не используйте имена доменов.
- Используйте звездочку с осторожностью. Она позволяет любому пользователю с допустимой учетной записью для сервера IBM SPSS Modeler (то есть в большинстве случаев всем пользователям в сети) входить в систему и изменять опции конфигурации.

Соединения сервера SPSS Modeler

Вы должны задать соединение с каждым сервером SPSS Modeler в вашей сети, которым собираетесь управлять. Затем надо зарегистрироваться на каждом сервере. Хотя соединение с сервером сохраняется

между сеансами Modeler Administration Console, регистрационные данные не сохраняются. Вы должны входить в систему каждый раз, когда запускаете Modeler Administration Console.

Чтобы задать соединение с сервером

1. Убедитесь, что служба сервера IBM SPSS Modeler запущена.
2. В меню Файл выберите **Создать** и затем **Соединение с управляемым сервером**.
3. На первой странице мастера введите имя для вашего соединения с сервером. Это имя предназначено для вас и должно быть описательным, например, *Production Server*. Убедитесь, что задан тип **Управляемый сервер IBM SPSS Modeler**, и нажмите кнопку **Далее**.
4. На второй странице введите имя хоста или IP-адрес сервера. Если вы измените порт по сравнению со значением по умолчанию, введите номер порта. Нажмите кнопку **Готово**. Новое соединение с сервером выводится на панели Администратор серверов.

Для выполнения административных задач надо войти в систему.

Чтобы войти в систему на сервере

1. На панели Администратор серверов щелкните дважды по серверу, на котором вы хотите войти в систему.
2. В диалоговом окне Вход в систему введите регистрационные данные. (Используйте вашу учетную запись пользователя для хоста сервера.) Нажмите кнопку **ОК**.

Если вход в систему завершается неудачно с сообщением **Невозможно получить права администратора на сервере**, обычно это вызвано неверной конфигурацией доступа администратора. Дополнительную информацию смотрите в разделе “Конфигурирование доступа с Modeler Administration Console” на стр. 23.

Если вход в систему завершается неудачно с сообщением **Не удалось соединиться с сервером '<сервер>'**, убедитесь, что ID пользователя и пароль заданы правильно, затем проверьте, работает ли служба сервера IBM SPSS Modeler. Например, в системе Windows перейдите на Панель управления > Администрирование > Службы и проверьте запись сервера IBM SPSS Modeler. Если в столбце Состояние не написано **Работает**, выберите эту строку на экране и щелкните по ссылке **Запустить**, затем снова попытайтесь войти в систему.

Когда вы вошли в систему на вашем сервере IBM SPSS Modeler, под именем сервера выводятся две опции - Конфигурирование и Мониторинг. Щелкните дважды по одной из этих опций.

Конфигурирование сервера IBM SPSS Modeler

На панели Конфигурация выводятся опции конфигурации для сервера IBM SPSS Modeler. Используйте эту панель, чтобы изменить опции, как вам требуется. Нажмите кнопку **Сохранить** на панели инструментов, чтобы сохранить изменения. (Обратите внимание на то, что изменение номера порта требует перезапуска сервера, чтобы внесенные изменения вступили в силу.)

Опции описаны в последующих разделах с соответствующими строками файла *options.cfg* в скобках. Опции, доступные только в *options.cfg*, описаны в конце раздела.

Подключения/Сеансы

Максимальное число соединений. (*max_sessions*) Максимальное количество пользователей, подключенных одновременно. Значение -1 означает отсутствие предела.

Номер порта. (*port_number*) Номер порта, на котором сервер SPSS Modeler ожидает подключения. Измените его, если это значение по умолчанию уже использует другая программа. Конечные пользователи должны знать этот номер порта, чтобы использовать сервер SPSS Modeler.

Соединение Analytic Server (сервер аналитических служб)

URL сервера Analytic. (*as_url*) URL для Analytic Server (сервер аналитических служб), в том числе IP-адрес сервера, порт и арендатор, в который входит установка сервера SPSS Modeler.

Приглашение ввода пароля. (`as_prompt_for_password`) Задайте `N`, если сервер SPSS Modeler сконфигурирован с той системой аутентификации для пользователей и паролей, которую использует Analytic Server (сервер аналитических служб); иначе задайте `Y`.

Доступ к файлу данных

Ограничить доступ только каталогом файлов данных. (`data_files_restricted`) Если для этой опции задано значение `yes`, доступ к файлам данных ограничивается стандартным каталогом данных, который указан в поле **Путь файлов данных** ниже.

Путь файлов данных. (`data_file_path`) Список дополнительных каталогов, в которых клиентам разрешено выполнять чтение и запись файлов данных. Эта опция игнорируется, если не включена опция **Ограничить доступ только каталогом файлов данных**. Обратите внимание на то, что во всех путях надо использовать прямую дробную черту. Задайте несколько каталогов через точку с запятой (например, *[путь установки сервера]/data;c:/data;c:/temp*).

Ограничить доступ только каталогом программных файлов. (`program_files_restricted`) Если для этого параметра задано значение `yes`, доступ к программным файлам ограничивается стандартным каталогом `bin`, указанным ниже в поле **Путь программных файлов**.

Путь программных файлов. (`program_file_path`) Список дополнительных каталогов, из которых клиентам разрешено выполнять программы. Эта опция игнорируется, если не включена опция **Ограничить доступ только каталогом программных файлов**. Обратите внимание на то, что во всех путях надо использовать прямую дробную черту. Задайте несколько каталогов через точку с запятой.

Максимальный размер файлов (Мбайт). (`max_file_size`) Максимальный размер временных и экспортированных файлов данных, созданных при выполнении потока (не применяется к наборам данных SAS и SPSS Statistics). Значение `-1` означает отсутствие предела.

Временный каталог. (`temp_directory`) Каталог, используемый для хранения временных файлов данных (файлов кэша). В идеале этот каталог должен находиться на отдельном высокоскоростном диске или контроллере, так как скорость доступа к этому каталогу может иметь решающее значение для производительности. Вы можете задать несколько временных каталогов, разделяя их запятыми. Они должны быть расположены на различных дисках; первый каталог используется чаще всего, а дополнительные каталоги служат для хранения временных рабочих файлов, когда определенные операции подготовки данных (такие как сортировка) используют параллелизм при выполнении. Если разрешить каждому потоку выполнения использовать отдельные диски для временного хранения, это может повысить производительность. Используйте во всех спецификациях путей прямую дробную черту. *Примечание:* Временные файлы генерируются в этом каталоге при запуске сервера SPSS Modeler. Убедитесь, что у вас есть необходимые права доступа к этому каталогу (например, если временный каталог находится в общей сетевой папке), иначе запуск сервера SPSS Modeler завершится неудачно.

Путь exe-файлов Python . (`python_exe_path`) Полный путь к исполняемому файлу Python, включая имя исполняемого файла. Заметим, что для опции **Ограничить доступ только каталогом программных файлов** может потребоваться задать значение `No` в зависимости от положения установки `python`.

Производительность/Оптимизация

Перезапись потока. (`stream_rewriting_enabled`) Позволяет серверу оптимизировать потоки, перезаписывая их. Например, сервер может перенести операции сокращения данных ближе к узлу источника, чтобы сократить размер набора данных по возможности раньше. Отключение этой опции обычно рекомендуется, только если оптимизация приводит к ошибкам или другим неожиданным результатам. Этот параметр переопределяет соответствующее значение оптимизации клиента.

Параллелизм. (`max_parallelism`) Описывает число параллельных потоков обработки, которые разрешается запускать SPSS Modeler при выполнении потока. Задание значения 0 или любого отрицательного числа указывает IBM SPSS Modeler запускать столько потоков, сколько доступно процессоров на компьютере; значение по умолчанию для этой опции `-1`. Чтобы отключить параллельную обработку (для компьютеров

с несколькими процессорами), задайте для этой опции значение 1. Чтобы разрешить параллельную обработку с ограничением числа потоков, задайте значение меньше числа процессоров на вашем компьютере. Заметим, что гиперпоточный или двухъядерный процессор считается как два процессора.

Размер буфера (в байтах). (`io_buffer_size`) Файлы данных, передаваемые с сервера на клиент, проходят через буфер указанного размера.

Сжатие кэша. (`cache_compression`) Целое значение от 0 до 9, которое управляет сжатием кэша и прочих файлов во временном каталоге сервера. Сжатие сокращает объем используемого места на диске, что может быть важным при ограниченном пространстве, и повышает производительность путем сокращения объема дисковых операций по созданию и чтению кэша. При сжатии увеличивается процессорное время, но это почти всегда компенсируется сокращением времени доступа к диску. Обратите внимание на то, что сжать можно только определенные кэши - те, обращение к которым выполняется последовательно. Эта опция не применяется к кэшам произвольного доступа, например, к таким, которые используются алгоритмами обучения сети. Значение 0 полностью отключает сжатие. Значения от 1 и выше обеспечивают повышение степени сжатия, однако при этом растет и время доступа. Значение по умолчанию - 1; более высокие значения нужны редко, разве что для вас критично место на диске.

Множитель использования памяти. (`memory_usage`) Управляет долей физической памяти, отводимой для сортировки и другого кэширования в памяти. Значение по умолчанию - 100, что соответствует примерно 10% физической памяти. Увеличение этого значения повышает производительность сортировки, если доступна свободная память, однако будьте осторожны, увеличивая его, так как это может привести к избыточной подкачке страниц.

Предельный процент памяти моделирования. (`modelling_memory_limit_percentage`) Управляет долей физической памяти, выделяемой для обучения моделей нейросети, Коонена и k -средних. Значение по умолчанию - 25%. Увеличение этого значения повышает производительность обучения, если доступна свободная память, однако будьте осторожны, увеличивая его, так как это может привести к избыточной подкачке при сбросе страниц на диск.

Разрешить переопределение памяти моделирования. (`allow_modelling_memory_override`) Включает или отключает опцию **Оптимизировать для скорости** для некоторых узлов моделирования. Значение по умолчанию - включено. Эта опция позволяет алгоритму моделирования запрашивать всю доступную память, обходя опцию предельного процента. Вы можете захотеть отключить эту опцию, если требуется использовать ресурсы памяти на компьютере сервера совместно.

Максимальный и минимальный порты сервера. (`max_server_port` и `min_server_port`) Задают диапазон номеров портов, которые можно использовать для дополнительных соединений гнезд между клиентом и сервером, требуемых для выполнения интерактивных моделей и потоков. При этом сервер должен ожидать соединений на другом порту; если не ограничивать диапазон, пользователи на системах с брандмауэрами могут столкнуться с проблемами. По умолчанию задается значение -1, что означает "без ограничений". Таким образом, например, чтобы сервер ожидал соединений на порту 8000 и последующих, надо задать для `min_server_port` значение 8000, а для `max_server_port` - значение -1.

Заметим, что вы должны открыть дополнительные порты помимо главного порта сервера, чтобы открыть или выполнить поток, и соответственно больше портов, если вы хотите открывать или выполнять дополнительные потоки. Это необходимо для получения обратной связи от выполнения потока.

По умолчанию IBM SPSS Modeler будет использовать любой доступный открытый порт; если он не находит ни одного (например, если они закрыты на брандмауэре), при выполнении потока выводится ошибка. Чтобы сконфигурировать диапазон портов, IBM SPSS Modeler будет нужно два доступных открытых порта (в дополнение к главному порту сервера) на одновременно выполняемый поток, плюс по дополнительному порту для каждого соединенного клиента, у которого есть соединение ODBC. Чтобы выполнять больше параллельных потоков, можно расширить диапазон номеров портов и открыть соответственно больше портов на брандмауэре.

Примечание: Если вы изменяете эти параметры, надо перезапустить сервер SPSS Modeler, чтобы изменение вступило в силу.

Оптимизация выборки массива. (`sql_row_array_size`) Управляет способом, которым сервер SPSS Modeler выбирает данные из источника данных ODBC. Значение по умолчанию - 1, то есть выборка по одной строке за раз. Увеличение этого значения указывает серверу читать информацию более крупными кусками, выполняя выборку указанного числа строк в массив. Для некоторых сочетаний операционная система/база данных это может дать повышение производительности для операторов SELECT.

SQL

Максимальная длина строки SQL. (`max_sql_string_length`) Для строки, импортируемой из базы данных с SQL, максимальное число символов, гарантированно передаваемых успешно. В зависимости от операционной системы более длинные строковые значения могут усекаться справа без предупреждения. Допустимый диапазон - от 1 до 65535 символов. Это свойство применяется также к узлу Экспорт баз данных.

Автоматическое генерирование SQL. (`sql_generation_enabled`) Разрешает автоматическое генерирование SQL для потоков, что может существенно повысить производительность. Значение по умолчанию - включено. Отключение этой опции рекомендуется только в том случае, когда база данных не может поддерживать запросы, которые передает сервер SPSS Modeler. Обратите внимание на то, что это значение переопределяет соответствующий параметр оптимизации клиента, а также на то, что для целей скоринга генерирование SQL надо включить отдельно для каждого узла моделирования независимо от значения данного параметра.

Длина строки SQL по умолчанию. (`default_sql_string_length`). Задаёт ширину строчных столбцов по умолчанию, которые будут создаваться в таблицах кэша базы данных. Строчные поля в таблицах кэширования баз данных будут создаваться с шириной по умолчанию 255, если нет информации о типе восходящего потока. Если в ваших данных есть более длинные значения, либо реализуйте узел типа восходящего потока с этими значениями, либо задайте для этого параметра достаточно большое значение, соответствующее этим строчным значениям данных.

Поддержка записи в журнал типа SQL. (`supported_sql_type_logging`). Только для целей отладки; не изменяйте это значение без указания представителя технической поддержки.

Включить пользовательские функции базы данных. (`db_udf_enabled`). Если задано значение Y (по умолчанию), опция генерирования SQL порождает пользовательские функции SQL вместо чистого кода SQL SPSS Modeler. Пользовательские функции SQL обычно производительнее чистого SQL.

SSL

Включить SSL. (`ssl_enabled`) Включает шифрование SSL для соединений между SPSS Modeler и сервером SPSS Modeler.

Файл сертификата. (`ssl_certificate_file`) Файл общедоступного сертификата SSL для загрузки при запуске сервера.

Файл секретного ключа. (`ssl_private_key_file`) Файл секретного ключа SSL для загрузки при запуске сервера.

Пароль секретного ключа. (`ssl_private_key_password`) Ключ пароля для заданного сертификата. Обратите внимание на то, что этот ключ должен быть подписан хотя бы одним сертификатом из перечисленных в сконфигурированном файле сертификата.

Координатор конфигурирования процессов

Хост. (`cor_host`) Имя хоста или IP-адрес службы координатора процессов. Значение по умолчанию “spsscscop” - это общее имя, которой администраторы могут добавить как алиас для хоста IBM SPSS Collaboration and Deployment Services на DNS.

Номер порта. (cop_port_number) Номер порта службы координатора процессов. Значение по умолчанию для IBM SPSS Collaboration and Deployment Services - 8080.

Корень контекста. (cop_context_root) URL службы координатора процессов.

Регистрационное имя. (cop_user_name) Имя пользователя для аутентификации на службе координатора процессов. Это регистрационное имя IBM SPSS Collaboration and Deployment Services, так что оно может содержать добавленный защитой префикс (например, ad/jsmith).

Пароль. (cop_password) Пароль для аутентификации на службе координатора процессов.

Включен. (cop_enabled) Определяет, должен ли сервер пытаться регистрироваться на координаторе процессов. Значение по умолчанию - *не* регистрироваться, так как администратор должен выбрать, какие службы сделать видимыми координатору процессов.

Имя сервера. (cop_service_name) Имя данного экземпляра сервера SPSS Modeler; по умолчанию это имя хоста.

Описание. (cop_service_description) Описание этого экземпляра.

Интервал обновления (минут). (cop_update_interval) Интервал между сообщениями работоспособности в минутах; значение по умолчанию - 2.

Вес. (cop_service_weight) Вес данного экземпляра, целое число от 1 до 10. Чем больше вес, тем больше соединений будут использоваться. Значение по умолчанию - 1.

Хост сервера. (cop_service_host) Полное имя хоста сервера IBM SPSS Modeler. Имя хоста по умолчанию строится автоматически; администратор может переопределить это значение для хостов распределенного базирования.

Опции, видимые в файле options.cfg

administrators. Задайте имена пользователей, которым вы хотите предоставить доступ администратора. Дополнительную информацию смотрите в разделе “Конфигурирование доступа с Modeler Administration Console” на стр. 23.

allow_config_custom_overrides. Не изменяйте это значение без указания представителя технической поддержки.

fips_encryption. Разрешает FIPS-совместимое шифрование; по умолчанию - N.

max_transfer_size. Только для внутреннего использования, не изменяйте это значение.

shell. (Только для серверов UNIX) Переопределяет значение по умолчанию для оболочки UNIX, например, shell, "/usr/bin/ksh". По умолчанию IBM SPSS Modeler использует оболочку, определенную в профиле пользователя, соединяющегося с сервером IBM SPSS Modeler.

start_process_as_login_user. Задайте для этого параметра значение Y, если вы используете единую регистрацию и соединены с сервером IBM SPSS Collaboration and Deployment Services, на котором применяются способ аутентификации, отличный от хоста сервера IBM SPSS Modeler. Задайте значение Y также, если вы запускаете сервер SPSS Modeler с базой данных секретных паролей.

time_encode_execution_log. Задайте значение Y, если вы хотите включать в файл *message.log* (в папке *log* каталога установки сервера SPSS Modeler) включать в себя отдельные времена выполнения для каждого узла в потоке.

`use_bigint_for_count`. Когда число подсчитываемых записей больше, чем может быть записано в обычное целое (2³²-1), задайте для этой опции значение Y. Когда для этой опции задано значение Y и поток соединен с DB2 или с SQL Server, для подсчета числа записей будет использоваться COUNT_BIG() (например, для поля `Record_Count`, генерируемого узлом агрегации).

Мониторинг сервер SPSS Modeler

На панели мониторинга Modeler Administration Console выводится снимок всех процессов, выполняемых на компьютере сервера SPSS Modeler, подобно менеджеру задач Windows. Чтобы активировать панель мониторинга, щелкните дважды по узлу Мониторинг у требуемого сервера на панели Администратор серверов. При этом на панель будет выведен текущий снимок данных от сервера. Эти данные обновляются с показанной частотой (по умолчанию раз в минуту). Чтобы обновить данные вручную, нажмите кнопку **Обновить**. Чтобы вывести в этом списке только процессы сервера SPSS Modeler, нажмите кнопку **Отфильтровать процессы не с сервера SPSS Modeler**.

Использование файла options.cfg

Файл `options.cfg` находится в каталоге *[путь установки сервера]/config*. Каждый параметр представлен парой разделенных запятой значений, где **имя** - название опции, а **значение** - значение этой опции. Комментарии отмечены знаком решетки (#).

Примечание: Большинство опций конфигурации можно изменить при помощи IBM SPSS Modeler Administration Console, а не через этот файл конфигурации, хотя есть несколько исключений. Дополнительную информацию смотрите в разделе “Опции, видимые в файле options.cfg” на стр. 28.

IBM SPSS Modeler Administration Console позволяет избежать перезапусков сервера для всех опций, кроме порта сервера. Дополнительную информацию смотрите в разделе “Использование IBM SPSS Modeler Administration Console” на стр. 23.

Опции конфигурации, которые могут быть добавлены в файл, используемый по умолчанию

По умолчанию кэширование в базе данных на сервере IBM SPSS Modeler включено. Эту возможность можно отключить, добавив следующую строку в файл `options.cfg`:

```
enable_database_caching, N
```

В результате временные файлы будут создаваться на сервере, а не в базе данных.

Чтобы просмотреть или изменить опции конфигурации сервера IBM SPSS Modeler:

1. Откройте файл `options.cfg` в текстовом редакторе.
2. Найдите интересующие вас опции. Полный список опций смотрите в разделе “Конфигурирование сервера IBM SPSS Modeler” на стр. 24.
3. Измените значения нужным образом. Учтите, что во всех путях необходимо использовать прямую дробную черту (/), а не обратную, как в разделителях путей.
4. Сохраните файл.
5. Остановите и перезапустите сервер IBM SPSS Modeler, чтобы изменения вступили в силу. Дополнительную информацию смотрите в разделе “Запуск и остановка сервера IBM SPSS Modeler” на стр. 21.

Закрытие неиспользуемых соединений с базой данных

По умолчанию в IBM SPSS Modeler кэшируется по крайней мере одно соединение с базой данных после обращения к этому соединению. Сеанс базы данных остается открытым, даже когда потоки, требующие доступа к базе данных, не выполняются.

Кэширование соединений с базой данных может улучшить время выполнения, устраняя потребность в повторном соединении IBM SPSS Modeler с базой данных при каждом выполнении потока. Однако в

некоторых средах для прикладных программ важно высвободить ресурсы базы данных как можно быстрее. Если слишком много сеансов IBM SPSS Modeler поддерживают более не используемые соединения с базой данных, ресурсы базы данных могут исчерпаться.

Этой возможности можно избежать, выключив опцию IBM SPSS Modeler `cache_connection` в пользовательском файле конфигурации базы данных. Это может придать IBM SPSS Modeler дополнительную устойчивость к отказам в соединении с базой данных (например, к истечению срока ожидания), который возможен при длительном использовании соединений сеансом IBM SPSS Modeler.

Чтобы заставить закрыть неиспользуемые соединения с базой данных:

1. Найдите каталог *[путь установки сервера]/config*.
2. Добавьте следующий файл (или откройте его, если он уже существует):
odbc-custom-properties.cfg
3. Добавьте в этот файл строку:
`cache_connection, N`
4. Сохраните и закройте файл.
5. Перезапустите сервер IBM SPSS Modeler, чтобы изменения вступили в силу.

Использование SSL для защищенной передачи данных

Secure Sockets Layer (SSL) - это протокол для шифрования данных, передаваемых между двумя компьютерами. SSL обеспечивает безопасный обмен данными между компьютерами. SSL может зашифровать аутентификацию имени пользователя/пароля и содержание обмена информацией между сервером и клиентом.

Как работает SSL

SSL основана на открытых и секретных ключах сервера, в дополнение к сертификату открытого ключа, который связывает идентичность сервера с его открытым ключом.

1. Когда клиент соединяется с сервером, клиент аутентифицирует сервер, используя сертификат открытого ключа.
2. Затем клиент генерирует случайное число, шифрует число открытым ключом сервера и возвращает зашифрованное сообщение на сервер.
3. Сервер дешифрует это случайное число своим секретным ключом.
4. На основе этого случайного числа сервер и клиент создают ключи сеанса, используемые для шифрования и дешифрования последующей информации.

Сертификат открытого ключа обычно подписывается центром сертификации. Центры сертификации, такие как VeriSign и Thawte, - это организации, которые выпускают и аутентифицируют параметры защиты, содержащиеся в сертификатах открытого ключа, и управляют ими. По существу центр сертификации подтверждает идентификационные данные сервера. Центр сертификации обычно взимает плату за сертификат, но можно сгенерировать и самоподписанные сертификаты.

Обеспечение защищенной связи клиент-сервер и сервер-сервер с использованием SSL

Основные шаги в обеспечении безопасной связи клиент-сервер и сервер-сервер с использованием SSL:

1. Установите OpenSSL на компьютере сервера.
2. Получите и установите сертификат и ключи SSL.
3. Включите и сконфигурируйте SSL в прикладной программе администратора сервера (IBM SPSS Collaboration and Deployment Services Deployment Manager).
4. При необходимости установите шифрование неограниченной силы на клиентских компьютерах.

5. Добавьте сертификат на склад ключей клиента.
6. Дайте пользователям указание включать SSL при соединении с сервером.

Примечание: Иногда продукт сервера может работать как клиент. В примере рассматривается сервер IBM SPSS Statistics, соединяющийся с IBM SPSS Collaboration and Deployment Services Repository. В этом случае сервер IBM SPSS Statistics выступает как *клиент*.

Установите OpenSSL

Если OpenSSL более недоступен на сервере, необходимо установить его.

1. Скачайте OpenSSL с сайта <http://www.openssl.org/>. Проверьте, что версия OpenSSL соответствует версии сервера:

Таблица 1. Версии OpenSSL.

Продукт сервера	Совместимая версия OpenSSL
IBM SPSS Statistics 20-22	1.0.0
IBM SPSS Statistics 17-19 (<i>не Linux[®] on System z[®]</i>) IBM SPSS Modeler 16 (<i>не Linux[®] on System z[®]</i>)	0.9.8 и ее подверсии (0.9.8a, 0.9.8b и так далее)
IBM SPSS Statistics 19 (Linux [®] on System z [®]) IBM SPSS Modeler 16 (Linux [®] on System z [®])	1.0.0

Примечание: сервер IBM SPSS Modeler для UNIX, конфигурируемый для SSL, ищет ссылку на две библиотеки с именами вида (.so может быть специфичным для платформы):

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

При использовании подверсий SSL (0.9.8a, 0.9.8b, и так далее), необходимо создать символические ссылки для этих библиотек с версиями *xxx.0.9.8*, которых ожидает сервер IBM SPSS Modeler.

2. Следуйте инструкциям по установке и конфигурированию программного обеспечения. Если вы выполнили сборку OpenSSL самостоятельно, имейте в виду следующие правила:

Windows. OpenSSL должен быть собран с библиотеками DLL (которые по умолчанию являются многопоточными).

UNIX. OpenSSL должен поддерживать несколько потоков (что не всегда задано по умолчанию) и совместно используемые библиотеки.

3. Убедитесь, что система может найти библиотеки SSL:

Windows. Убедитесь, что модули OpenSSL включены в путь системы.

UNIX. Добавьте библиотеки SSL в конец пути к библиотеке загрузки, заданного в файле *modelersrv.sh*.

Например, в Solaris добавьте в сценарий запуска сервера следующую строку:

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

Получение и установка сертификата и ключей SSL

1. Получите сертификат SSL и файл ключа. Это можно сделать двумя способами:

- Приобретите их в общедоступном центре сертификации (например, VeriSign или Thawte). Общедоступный центр сертификации подписывает сертификат для проверки сервера, который его использует.
- Сгенерируйте ключ и файлы сертификата с помощью внутреннего центра выдачи самоподписанных сертификатов. Для этого воспользуйтесь имеющимся в OpenSSL инструментом управления сертификатами или найдите в Интернете инструкции по созданию самоподписанного сертификата SSL.

2. Скопируйте сертификат и один или несколько файлов ключа в локальный каталог или каталоги или на сервер. Общедоступный и секретный ключ можно хранить в отдельных каталогах. Можно также записать их в один файл. Нужно, чтобы права доступа не допускали случайных просмотров каталога, который содержит секретный ключ.

Включите и сконфигурируйте SSL в IBM SPSS Collaboration and Deployment Services Deployment Manager

1. Запустите прикладную программу администратора сервера (IBM SPSS Collaboration and Deployment Services Deployment Manager) и подключитесь к серверу.
2. На странице конфигурации задайте для **Secure Sockets Layer** значение Да.
3. В разделе **Файл открытого ключа SSL** задайте полный путь к файлу открытого ключа.
4. В разделе **Файл секретного ключа SSL** задайте полный путь к файлу секретного ключа.
Примечание: Если открытый и секретный ключи хранятся в одном файле, укажите в разделах **Файл открытого ключа SSL** и **Файл секретного ключа SSL** один и тот же файл.
5. Выберите в меню:
Файл > Сохранить
6. Перезапустите службу сервера или демон. При перезапуске вас попросят ввести пароль SSL. В Windows можно выбрать **Запомнить этот пароль** для безопасного хранения пароля. Эта опция избавляет от необходимости вводить пароль каждый раз, когда сервер запущен.

Конфигурирование SSL на сервере IBM SPSS Modeler

Если для доступа к секретному файлу ключа SSL требуется пароль, способ его предоставления зависит от операционной системы.

Системы Windows

В системах Windows предоставлять пароль можно одним из следующих способов:

- Задать его в файле *options.cfg* при помощи опции **ssl_private_key_password**.
- Задать его в окне командной строки. Если опции **ssl_private_key_password** нет в *options.cfg*, при запуске появится окно командной строки.

Системы UNIX

В системах UNIX предоставлять пароль можно одним из следующих способов:

- Задать его в файле *options.cfg* при помощи опции **ssl_private_key_password**.
- Добавить ключ в файл *modelersrv.sh* для запроса пароля при запуске.

При использовании последней опции найдите в файле *modelersrv.sh* следующую строку:

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

Добавьте ключ `-request_ssl_password`:

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

Установка шифрования неограниченной сложности

В среде выполнения Java, поставляемой с продуктом, включено шифрование экспортного варианта США. Для лучшей защиты ваших данных рекомендуется выполнить обновление до шифрования неограниченной силы.

IBM J9

1. Скачайте файлы Java Cryptography Extension (JCE) Unlimited Strength Jurisdiction Policy Files для вашей версии SDK с сайта IBM.com.

2. Извлеките файлы Unlimited Jurisdiction Policy, упакованные в сжатом файле. Сжатый файл содержит файлы *US_export_policy.jar* и *local_policy.jar*. В установке WebSphere Application Server перейдите в каталог *\$JAVA_HOME/jre/lib/security* и создайте резервные копии файлов *US_export_policy.jar* и *local_policy.jar*.
3. Замените ваши файлы *US_export_policy.jar* и *local_policy.jar* на файлы, которые скачали.
4. Включите защиту на консоли администрирования WebSphere Application Server. Убедитесь, что все агенты узла в ячейке заранее активированы. Дополнительную информацию смотрите в документации по WebSphere. Учтите, что вам нужно выбрать доступное определение пространства из списка **Защита > Защита администрирования, программ и инфраструктуры** и затем нажать кнопку **Задать в качестве текущего**, чтобы защиты была включена после перезапуска сервера.
5. Закройте консоль администрирования.
6. Остановите сервер.
7. Перезапустите сервер.

Sun Java

1. Скачайте файлы Java Cryptography Extension (JCE) Unlimited Strength Jurisdiction Policy Files для вашей версии SDK с сайта Sun Java.
2. Извлеките скачанный файл.
3. Скопируйте два файла *.jar*: *local_policy.jar* и *US_export_policy.jar* в положение *<папка_установки>/jre/lib/security*, где *<папка_установки>* - папка, в которую установлен продукт.

Инструментирование пользователей: как включить SSL

Когда пользователи соединяются с сервером через клиентский продукт, им нужно для соединения с сервером включить SSL в диалоговом окне. Обязательно укажите пользователям правильный переключатель для этого. Например, в IBM SPSS Modeler пользователям нужно включить переключатель **Обеспечить защищенное соединение** для репозитория, для чего в меню **Инструменты** надо выбрать опцию **Репозиторий**, чтобы открыть диалоговое окно Репозиторий: Сервер.

Глава 5. Обзор производительности

На реальную производительность при анализе данных влияет много факторов, от сервера и конфигурации базы данных до порядка отдельных узлов в потоке. В целом наилучшей производительности можно достичь следующим образом:

- Храните данные в СУБД и используйте генерирование и оптимизацию SQL, где только возможно.
- Используйте аппаратные средства, соответствующие критериям, указанным в Глава 2, “Рекомендации по архитектуре и аппаратным средствам”, на стр. 7, или превосходящие их.
- Убедитесь, что параметры производительности клиента и сервера сконфигурированы правильно. Обратите внимание на то, что если SPSS Modeler соединен с установкой сервера SPSS Modeler, параметры производительности и оптимизации сервера переопределяют соответствующие параметры на клиенте.
- Разработайте потоки для максимальной производительности.

Дополнительную информацию о каждом из этих факторов производительности смотрите в следующих разделах.

Параметры производительности и оптимизации сервера

Некоторые параметры сервера IBM SPSS Modeler можно сконфигурировать для оптимизации производительности. Эти параметры можно настроить с помощью прикладной программы IBM SPSS Modeler Administration Console, поставляемой с сервером IBM SPSS Modeler. Дополнительную информацию смотрите в разделе “Использование IBM SPSS Modeler Administration Console” на стр. 23.

Параметры сгруппированы под заголовком **Производительность и оптимизация** в окне конфигурации IBM SPSS Modeler Administration Console. Эти параметры предварительно сконфигурированы для оптимальной производительности в большинстве установок. Однако вам, возможно, понадобится скорректировать их в зависимости от ваших аппаратных средств, размера наборов данных и содержимого ваших потоков. Дополнительную информацию смотрите в разделе “Производительность/Оптимизация” на стр. 25.

Параметры производительности и оптимизации клиента

Параметры производительности и оптимизации клиента доступны на вкладке Опции диалогового окна Свойства потока. Для вывода этих опций выберите в меню клиента:

Инструменты > Свойства потока > Опции > Оптимизация

С помощью параметров оптимизации можно оптимизировать производительность потока. Обратите внимание на то, что параметры производительности и оптимизации на сервер IBM SPSS Modeler (если они используются) переопределяют аналогичные параметры на клиенте.

Примечание: Моделирование баз данных и оптимизация SQL требуют, чтобы на компьютере IBM SPSS Modeler была включена возможность соединения с сервером IBM SPSS Modeler. При включенной возможности соединения можно обращаться к алгоритмам баз данных, выполнять обратный перенос SQL непосредственно с клиента сервером IBM SPSS Modeler и обращаться к серверу IBM SPSS Modeler. Проверьте текущее состояние лицензии, для чего в меню клиента сервером IBM SPSS Modeler выберите:

Справка > О программе > Дополнительные подробности

Если возможность соединения включена, на вкладке Состояние лицензии вы увидите опцию **Разрешение для сервера**.

Дополнительную информацию смотрите в разделе “Соединение с сервер IBM SPSS Modeler” на стр. 15.

Примечание: Поддержка обратного переноса SQL и оптимизации зависит от типа используемой базы данных. Наиболее свежие сведения о базах данных и драйверах ODBC, поддерживаемых и проверенных для использования с IBM SPSS Modeler 16, смотрите на сайте технической поддержки <http://www.ibm.com/support>.

Включить перезапись в потоке. Выберите эту опцию, чтобы разрешить перезапись потока в IBM SPSS Modeler. Доступны четыре типа перезаписи, можно выбрать один или оба. Перезапись потока изменяет порядок узлов в потоке в фоновом режиме, обеспечивая более эффективную работу без изменения семантики потока.

- **Оптимизировать построение SQL.** Эта опция позволяет изменять порядок узлов в потоке, чтобы передавать больше операций при помощи генерирования SQL для выполнения в базу данных. При обнаружении узла, который не может быть переведен в SQL, оптимизатор будет искать далее нисходящие узлы, которые можно перевести в SQL и безопасно переставить в положение перед проблемным узлом, не влияя на семантику потока. База данных может выполнять операции эффективнее, чем IBM SPSS Modeler, а кроме того, такие переносы помогают сократить размер набора данных, возвращаемого в IBM SPSS Modeler для обработки. Это, в свою очередь, может сократить сетевой трафик и ускорить операции потока. Обратите внимание на то, что оптимизация SQL не окажет влияния, если не включить переключатель **Генерировать SQL**.
- **Оптимизировать выражение CLEM.** Эта опция позволяет оптимизатору искать выражения CLEM, которые можно предварительно обработать до запуска потока, чтобы увеличить скорость работы. Простой пример - если у вас есть выражение типа $\log(\text{salary})$, оптимизатор вычислит фактическое значение salary и передаст его на обработку. Это можно использовать, чтобы улучшить обратный перенос SQL и повысить производительность сервера IBM SPSS Modeler.
- **Оптимизировать выполнение синтаксиса.** Этот метод перезаписи потока повышает эффективность операций, в которых участвует несколько узлов, содержащих синтаксис IBM SPSS Statistics. Оптимизация достигается путем объединения команд синтаксиса в одну операцию вместо выполнения каждой операции по отдельности.
- **Оптимизировать другое выполнение.** Этот метод перезаписи потока повышает эффективность операций, которые не могут быть делегированы базе данных. Оптимизация достигается путем максимально раннего сокращения объема данных в потоке. Поток перезаписывается для перестановки операций ближе к источнику данных при одновременном поддержании целостности данных, уменьшая таким образом нисходящий поток данных для таких дорогостоящих операций, как объединения.

Включить параллельную обработку. При работе на компьютере с несколькими процессорами эта опция позволяет системе распределять нагрузку по этим процессорам, что позволяет повысить производительность. Параллельная обработка может быть выгодной при использовании нескольких узлов или отдельных указанных ниже узлов: C5.0, Слияние (по ключевому слову), Сортировка, Интервал (ранговый и мозаичный методы) и Агрегирование (с использованием одного или нескольких ключевых полей).

Генерировать SQL. Выберите эту опцию, чтобы включить генерирование SQL, что позволит передать потоковые операции обратно в базу данных при помощи кода SQL и сгенерировать процессы выполнения, которые могут повысить производительность. Для дальнейшего повышения производительности можно выбрать **Оптимизировать генерацию SQL**, чтобы максимизировать число операций, переданных обратно в базу данных. Когда операции для узла будут переданы обратно в базу данных, узел после запуска потока будет выделен фиолетовым.

- **Кэширование баз данных.** Для потоков, генерирующих SQL, который будет выполняться в базе данных, данные могут попутно кэшироваться во временную таблицу в базе данных, а не в файловую систему. В сочетании с оптимизацией SQL это может привести к существенному выигрышу в производительности. Например, вывод потока, объединяющего несколько таблиц для создания представления исследования данных, может кэшироваться и при необходимости использоваться повторно. Если включено кэширование базы данных, просто щелкните правой кнопкой по любому неконечному узлу для кэширования данных в этой точке, и при следующем запуске потока кэш будет создан автоматически непосредственно в базе данных. Это позволяет генерировать SQL для нисходящих узлов, еще более повышая производительность. При необходимости эту опцию можно отключить, например, если правила

политики или разрешения не допускают записи данных в базу данных. Если кэширование базы данных или оптимизация SQL не разрешены, кэш будет записан в файловую систему.

- **Использовать упрощенное преобразование.** Эта опция включает преобразование данных из строк в числа или из чисел в строки, если они сохранены в подходящем формате. Например, если данные хранятся в базе данных в виде строки, но фактически содержат значимое число, они могут быть преобразованы для использования при обратном переносе.

Примечание: Ввиду небольших различий в реализации SQL, потоки, запущенные в базе данных, могут вернуть результаты, несколько отличные от возвращаемых при запуске в IBM SPSS Modeler. По аналогичным причинам эти различия могут также варьировать в зависимости от поставщика базы данных.

Использование и оптимизация базы данных

Сервер баз данных. Если возможно, создайте выделенный экземпляр базы данных для исследования данных, чтобы на производственный сервер не влияли запросы IBM SPSS Modeler. Операторы SQL, сгенерированные IBM SPSS Modeler, могут потребовать значительных ресурсов, поскольку на компьютере сервера IBM SPSS Modeler в одной базе данных возможно выполнение SQL несколькими задачами.

Исследование в базе данных. Многие поставщики баз данных поставляют расширения анализа данных для своих продуктов. Эти расширения позволяют выполнять операции исследования данных (таким как построение моделей или скоринг) на сервере баз данных или на отдельном выделенном сервере. Функции IBM SPSS Modeler исследования в базе данных дополняют и расширяют возможности генерирования SQL, позволяя поставщикам выпускать собственные версии расширений базы данных. В некоторых случаях этот подход позволяет избежать потенциально высоких затрат на передачу данных между IBM SPSS Modeler и базой данных. Еще большие преимущества может дать кэширование базы данных. Дополнительную информацию смотрите в файле *DatabaseMiningGuide.pdf*, который находится в папке *Documentation* на компакт-диске установки IBM SPSS Modeler.

Оптимизация SQL

Для лучшей производительности необходимо всегда стремиться максимизировать объем SQL, сгенерированного для использования производительности и масштабируемости базы данных. Только те части потока, которые нельзя скомпилировать в SQL, должен выполнять сервер IBM SPSS Modeler. Дополнительную информацию смотрите в разделе Глава 6, “Оптимизация SQL”, на стр. 39.

Закачка файлов данных

На данные, не хранящиеся в базе данных, не распространяются преимущества оптимизации SQL. Если данные, которые нужно проанализировать, еще не находятся в базе данных, можно загрузить их с помощью узла Вывод базы данных. Этот узел можно также использовать для хранения промежуточных наборов данных после подготовки данных, а также результатов внедрения.

IBM SPSS Modeler может взаимодействовать через интерфейс с внешними загрузчиками для многих общих систем баз данных. Вместе с программами поставляется несколько сценариев, которые доступны (с документацией) в подкаталоге */scripts* в папке установки IBM SPSS Modeler.

В следующей таблице показаны потенциальные выгоды от массовой загрузки для производительности. Данные показывают время, затраченное на экспорт 250000 записей и 21 поля в базу данных Oracle. Внешний загрузчик - утилита Oracle *sqlldr*.

Таблица 2. Выигрыш в производительности от массовой загрузки

Опция экспорта	Время (в секундах)
По умолчанию (ODBC)	409
Массовая загрузка через ODBC	52
Массовая загрузка через внешний загрузчик	33

Глава 6. Оптимизация SQL

Одна из наиболее мощных возможностей IBM SPSS Modeler - способность выполнять множество операций подготовки и исследования данных непосредственно в базе данных. Генерирование кода SQL, который можно передать обратно в базу данных для выполнения, позволяет выполнять многие операции, например, выборка, сортировка, создание производных полей и некоторые виды графических задач, в базе данных, а не на компьютере IBM SPSS Modeler или сервере IBM SPSS Modeler. Если вы работаете с большими наборами данных, эти **обратные переносы** могут существенно повысить производительность несколькими способами:

- Путем сокращения размера набора результатов, передаваемого из СУБД в IBM SPSS Modeler. Когда большие наборы результатов считываются через драйвер ODBC, сетевой ввод-вывод или драйвер могут работать недостаточно эффективно. Поэтому больше всего выигрывают от оптимизации SQL такие операции, как выбор строк и столбцов и агрегирование (узлы Выбор, Выборка, Агрегирование), что обычно уменьшает размер набора данных для передачи. Данные могут также кэшироваться во временной таблице базы данных в критических точках потока (например, после узлов Объединение и Выбор), что еще больше улучшает производительность.
- Путем использования производительности и масштабируемости базы данных. Рост эффективности связан с тем, что СУБД часто выигрывает от параллельной обработки, использования более мощных аппаратных средств, более сложного управления дисковым хранением и наличия индексов.

С учетом этих преимуществ, IBM SPSS Modeler разработан для максимизации количества SQL, сгенерированного каждым потоком, чтобы только те операции, которые нельзя скомпилировать в SQL, выполнялись сервером IBM SPSS Modeler. Однако из-за ограничений на выражения стандартного SQL (SQL 92) некоторые операции могут не поддерживаться. Дополнительную информацию смотрите в разделе “Советы по максимизации генерирования SQL” на стр. 43.

Примечание: Из-за небольших различий в реализации SQL потоки, запущенные в базе данных, могут вернуть несколько другие результаты, чем при выполнении в IBM SPSS Modeler. Эти различия могут также варьировать в зависимости от поставщика базы данных по сходным причинам. Например, в зависимости от конфигурации базы данных для чувствительности к регистру при сравнении и упорядочивании строк, потоки IBM SPSS Modeler, выполненные с использованием обратного переноса SQL, могут дать результаты, отличные от результатов без обратного переноса SQL. За консультацией по конфигурированию базы данных обратитесь к администратору базы данных. Для максимальной совместимости с IBM SPSS Modeler сравнения строк в базе данных должны быть чувствительными к регистру.

Примечание: Моделирование баз данных и оптимизация SQL требуют, чтобы на компьютере IBM SPSS Modeler была включена возможность соединения с сервером IBM SPSS Modeler. При включенной возможности соединения можно обращаться к алгоритмам баз данных, выполнять обратный перенос SQL непосредственно с клиента сервером IBM SPSS Modeler и обращаться к серверу IBM SPSS Modeler. Проверьте текущее состояние лицензии, для чего в меню клиента сервером IBM SPSS Modeler выберите:

Справка > О программе > Дополнительные подробности

Если возможность соединения включена, на вкладке Состояние лицензии вы увидите опцию **Разрешение для сервера**.

Дополнительную информацию смотрите в разделе “Соединение с сервером IBM SPSS Modeler” на стр. 15.

Примечание: Если потоки выполняются в базе данных Netezza, информация о дате и времени берется из этой базы данных. Она может отличаться от вашей локальной или заданной для сервера IBM SPSS Modeler даты и времени, если, например, база данных находится на компьютере, расположенном в другой стране или часовом поясе.

Требования к базе данных

Самую свежую информацию о базах данных и драйверах ODBC, работоспособность которых проверена и поддерживается при использовании с IBM SPSS Modeler, смотрите в матрице совместимости программного обеспечения на сайте технической поддержки (<http://www.ibm.com/support>).

Обратите внимание на то, что при помощи моделирования базы данных можно дополнительно улучшить производительность.

Как работает генерирование SQL

Начальные фрагменты потока, ведущего от исходных узлов базы данных, - основные объекты назначения для генерирования SQL. Когда попадает узел, который невозможно скомпилировать в SQL, данные извлекаются из базы данных и последующая обработка выполняется сервер IBM SPSS Modeler.

Во время подготовки потока и до выполнения процесс генерирования SQL происходит следующим образом:

- Сервер меняет порядок потоков, перемещая нисходящие узлы в "зону SQL", где этот процесс может выполняться безопасно. (Эту возможность на сервере можно отключить.)
- Выражения SQL строятся инкрементно, по мере продвижения от исходных узлов к конечным. В этой фазе процесс останавливается при обнаружении узла, который невозможно преобразовать в SQL или когда в SQL преобразовывается конечный узел (например, Таблица или Рисунок). Если у узла и его предшественников есть эквивалент SQL, в конце этой фазы каждый узел оказывается маркирован оператором SQL.
- Проверка допустимости SQL выполняется, начиная от узлов с наиболее сложными эквивалентами SQL, обратно к исходным узлам. Успешно проверенный SQL выбирается для выполнения.
- Узлы, для которых все операции сгенерировали SQL, выделяются фиолетовым на холсте потока. В зависимости от результатов вам может понадобиться далее реорганизовать поток в подходящих ситуациях, чтобы в полной мере воспользоваться выполнением в базе данных. Дополнительную информацию смотрите в разделе "Советы по максимизации генерирования SQL" на стр. 43.

Куда вносятся усовершенствования

Оптимизация SQL улучшает производительность ряда операций с данными:

- **Объединения (по ключевым словам).** Операции объединения могут улучшить оптимизацию в базах данных.
- **Агрегирование.** Узлы Агрегирование, Распределение и Web используют в своей работе агрегирование. Сводные данные требуют значительно меньшей пропускной способности, чем исходные данные.
- **Выбор.** Выбор записей на основе определенных критериев уменьшает количество записей.
- **Сортировка.** Сортировка записей - ресурсоемкое действие, которое эффективнее выполняется в базе данных.
- **Создание новых полей.** Новые поля эффективнее генерируются в базе данных.
- **Проекция полей.** сервер IBM SPSS Modeler извлекает только поля, которые требуются от базы данных для последующей обработки, сводя к минимуму требования к пропускной способности и памяти. То же самое справедливо в отношении избыточных полей в плоских файлах: несмотря на то, что сервер должен читать лишние поля, он не выделяет места для их хранения.
- **Скоринг.** SQL может генерироваться на основе деревьев решений, наборов правил, линейной регрессии и сгенерированных факторами моделей.

Пример генерирования SQL

Следующий поток присоединяет три таблицы базы данных при помощи ключевых операций и затем выполняет агрегирование и сортировку.

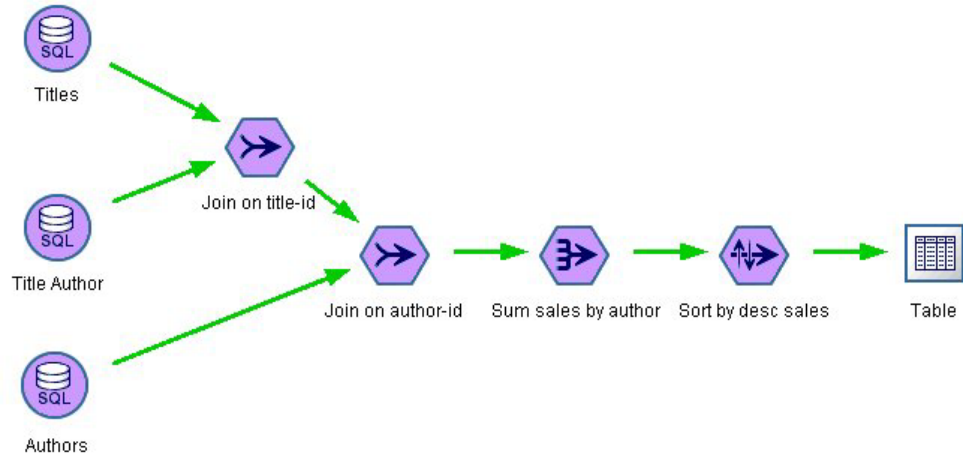


Рисунок 3. Оптимизированный поток с фиолетовыми узлами, указывающими на обратный перенос SQL (операции, выполняемые в базе данных)

Сгенерированный SQL

Сгенерированный SQL для этого потока:

```

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

```

Выполнение потока

Когда поток завершается узлом экспорта базы данных, для всего потока возможно выполнение в базе данных.

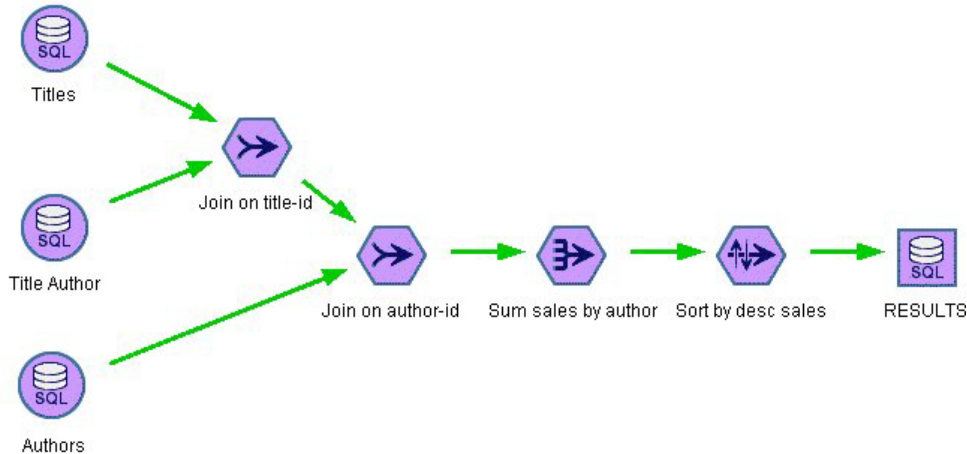


Рисунок 4. Весь поток, выполненный в базе данных

Конфигурирование оптимизации SQL

1. Установите драйвер ODBC и сконфигурируйте источник данных для базы данных, которую собираетесь использовать. Дополнительную информацию смотрите в разделе “Доступ к данным” на стр. 11.
2. Создайте поток, использующий исходный узел для получения данных из этой базы данных.
3. Проверьте, что генерирование SQL включено на клиенте и сервере, если это применимо. По умолчанию оно включено для обоих.

Чтобы разрешить оптимизацию SQL на клиенте

1. В меню Инструменты выберите **Свойства потока > Опции**.
2. Перейдите на вкладку Оптимизация. Выберите **Генерировать SQL**, чтобы разрешить оптимизацию SQL. Дополнительно можно выбрать другие параметры для повышения производительности. Дополнительную информацию смотрите в разделе “Параметры производительности и оптимизации клиента” на стр. 35.

Чтобы разрешить оптимизацию SQL на сервере

Поскольку настройки сервера переопределяют любые спецификации, сделанные на клиенте, должны быть включены обе опции конфигурации сервера: **Перезапись потока** и **Автоматическое генерирование SQL**. Дополнительную информацию о том, как изменить параметры сервера IBM SPSS Modeler, смотрите в разделе “Производительность/Оптимизация” на стр. 25

Чтобы разрешить оптимизацию при скоринге моделей

Для целей скоринга генерирование SQL должно быть включено отдельно для каждого узла моделирования, независимо от настроек уровня сервера или клиента. Это необходимо потому, что некоторые модели генерируют чрезвычайно сложные выражения SQL, которые нельзя эффективно оценить в базе данных. База данных может сообщить об ошибках при попытке выполнить сгенерированный SQL, из-за размера или сложности выражений SQL.

Чтобы определить, улучшает ли генерация SQL производительность для данной модели, может потребоваться некоторое количество проб и ошибок. Это можно сделать на вкладке Параметры после добавления сгенерированной модели в поток.

Предварительный просмотр сгенерированного SQL

Можно предварительно просмотреть сгенерированный SQL в журнале сообщений, прежде чем выполнить его в базе данных. Это может быть полезно для отладки и позволяет экспортировать сгенерированный SQL для последующего редактирования или запуска в базе данных. Предварительный просмотр также указывает, какие узлы будут переданы обратно в базу данных, что позволяет определить, можно ли переупорядочить поток для улучшения производительности.

1. Убедитесь, что в диалоговом окне Опции пользователя включены опции **Выводить SQL в журнале сообщений во время выполнения потока** и **Выводить подробности генерирования SQL в журнале сообщений во время подготовки потока**. Дополнительную информацию смотрите в разделе “Параметры производительности и оптимизации клиента” на стр. 35.
2. На холсте потока выберите узел или поток, который хотите предварительно просмотреть.
3. Нажмите кнопку **Предварительный просмотр SQL** на панели инструментов.
Все узлы, для которых сгенерирован SQL (и которые будут переданы обратно в базу данных после выполнения потока) на холсте потока будут окрашены в фиолетовый цвет.
4. Чтобы просмотреть сгенерированный SQL, выберите в меню:
Инструменты > Свойства потока > Сообщения...

Просмотр SQL для слепков моделей

В некоторых моделях можно сгенерировать SQL для слепка модели, передавая этап скоринга модели обратно в базу данных. Основное назначение этой функции - не повысить производительность, а разрешить потокам, содержащим эти слепки, передавать их полный SQL обратно. Дополнительную информацию смотрите в разделе “Узлы, поддерживающие генерирование SQL” на стр. 44.

Чтобы просмотреть SQL для слепка модели, поддерживающего генерирование SQL:

1. Выберите в слепке модели вкладку Параметры.
2. Выберите наиболее подходящую опцию, **Генерировать с поддержкой (без поддержки) пропущенных значений** или **Сгенерировать SQL для этой модели**.
3. В меню слепка модели выберите:
Файл > Экспортировать SQL
4. Сохраните файл.
5. Откройте файл для просмотра SQL.

Советы по максимизации генерирования SQL

Чтобы получить наибольший положительный эффект для производительности от оптимизации SQL, учтите следующее.

Порядок потоков. Генерирование SQL может быть остановлено, если у функции узла нет семантического эквивалента в SQL, поскольку функциональные возможности исследования данных в IBM SPSS Modeler богаче набора традиционных операций обработки данных, поддерживаемого стандартным SQL. В таких случаях генерирование SQL также подавляется для любых нисходящих узлов. Поэтому вы можете значительно улучшить производительность, переупорядочив узлы для помещения операций, останавливающих SQL, как можно дальше в нисходящем направлении. Оптимизатор SQL может сделать определенную работу по переупорядочению автоматически (просто убедитесь, что включена перезапись потока), но этот процесс можно усовершенствовать. Хорошо подходит для этого узел Выбор, который часто можно представлять. Дополнительную информацию смотрите в разделе “Узлы, поддерживающие генерирование SQL” на стр. 44.

Выражения CLEM. Если поток нельзя переупорядочить, можно изменить опции узла или выражения CLEM или иным образом видоизменить способ выполнения операции так, чтобы это больше не запрещало генерирование SQL. Узлы Извлечение, Выбор и подобные узлы обычно можно перевести в SQL, если у

операторов всех выражений CLEM есть эквиваленты SQL. Большинство операторов можно преобразовать, но существуют операторы, которые препятствуют генерированию SQL (особенно функции последовательностей [функции @]). Иногда генерирование прекращается из-за того, что сгенерированный запрос стал слишком сложным для обработки базой данных. Дополнительную информацию смотрите в разделе “Выражения и операторы CLEM, поддерживающие генерирование SQL” на стр. 48.

Несколько исходных узлов. В тех случаях, когда в потоке есть несколько исходных узлов базы данных, генерирование SQL независимо применяется к каждой ветви входного потока. Если генерирование останавливается на одной из ветвей, оно может продолжиться на другой. При объединении двух ветвей (если обе ветви можно выразить в SQL вплоть до слияния), само слияние часто можно заменить объединением в базе данных, и генерирование может продолжаться в нисходящем направлении.

Алгоритмы базы данных. Оценка модели всегда выполняется в сервер IBM SPSS Modeler, а не в базе данных, кроме случаев использования внутренних алгоритмов базы данных от Microsoft, IBM или Oracle.

Модели скоринга. Скоринг в базе данных поддерживается для некоторых моделей путем перевода сгенерированной модели в SQL. Однако некоторые модели генерируют чрезвычайно сложные выражения SQL, которые не всегда можно эффективно оценить в базе данных. По этой причине генерирование SQL необходимо включить отдельно для каждого узла модели. Если окажется, что узел модели препятствует генерированию SQL, перейдите на вкладку Параметры в диалоговом окне узла и выберите **Генерировать SQL для этой модели** (для некоторых моделей возможны дополнительные опции, управляющие генерированием). Проверьте с помощью тестов, что опция выгодна для вашей прикладной программы. Дополнительную информацию смотрите в разделе “Узлы, поддерживающие генерирование SQL”.

При тестировании узлов моделирования, чтобы увидеть, эффективно ли работает генерирование SQL, рекомендуется сначала сохранить все потоки от IBM SPSS Modeler. Некоторые системы базы данных могут зависнуть, пытаясь обработать (потенциально сложный) сгенерированный SQL, что требует закрытия IBM SPSS Modeler от диспетчера задач Windows.

Кэширование баз данных. Если для сохранения данных в критических точках потока (например, после узла объединения или агрегирования), используется кэш узла, убедитесь, что кэширование базы данных включено вместе с оптимизацией SQL. Это позволит данным кэшироваться в большинстве случаев во временной таблице в базе данных (а не в файловой системе). Дополнительную информацию смотрите в разделе “Конфигурирование оптимизации SQL” на стр. 42.

SQL, определяемый поставщиком. Большая часть сгенерированного SQL отвечает стандартам (SQL 92), но в некоторых случаях, когда это полезно, используются также некоторые нестандартные, специфичные для поставщика функции. Степень оптимизации SQL может быть различной в зависимости от источника базы данных.

Узлы, поддерживающие генерирование SQL

В следующие таблицы показаны узлы, представляющие операции исследования данных, которые поддерживают генерирование SQL. За исключением узлов моделирования базы данных, если узел не указан в этих таблицах, он не поддерживает генерирование SQL.

Можно предварительно просмотреть сгенерированный SQL, прежде чем выполнить его. Дополнительную информацию смотрите в разделе “Предварительный просмотр сгенерированного SQL” на стр. 43.

Таблица 3. Источники

Узел, поддерживающий генерирование SQL	Примечания
База данных	Этот узел используется для определения таблиц и представлений, предназначенных для дальнейшего анализа. На нем можно вводить операторы SQL. Избегайте наборов результатов с повторяющимися именами столбцов. Дополнительную информацию смотрите в разделе “Написание запросов SQL” на стр. 51.

Таблица 4. Операции записи

Узел, поддерживающий генерирование SQL	Примечания
Выбор	Поддерживает генерирование, только если поддерживается генерирование SQL для самого выражения выбора (смотрите выражения ниже). Если какие-то поля пусты, генерирование SQL не даст тех же результатов для отбрасывания, которые получаются в собственном IBM SPSS Modeler.
Выборка	Простая выборка поддерживает генерирование SQL, что может в разной степени зависеть от базы данных. Смотрите раздел Табл. 5.
Агрегирование	Поддержка генерирования SQL агрегации зависит от типа хранения данных. Смотрите раздел Табл. 6 на стр. 46.
Агрегирование RFM	Поддерживает генерирование, за исключением случаев сохранения даты второй или третьей по счету от конца транзакции или если содержит только недавние транзакции. Однако включение недавних транзакций оказывает эффект, если функция <code>datetime_date(YEAR,MONTH,DAY)</code> передана обратно.
Сортировка	
Слияние	<p>Для объединения по порядку SQL не генерируется.</p> <p>Объединение по ключевому слову с полным или частичным внешним объединением поддерживается, только если его поддерживает база данных или драйвер. Несовпадающие поля ввода можно переименовать при помощи узла Фильтр или вкладки Фильтр исходного узла.</p> <p>Поддерживает генерирование SQL для объединения по условию.</p> <p>Для всех типов объединения <code>SQL_SP_EXISTS</code> не поддерживается, если входные данные происходят из разных баз данных.</p>
Добавление в конец	Поддерживает генерирование, если входные данные не отсортированы.
Особый тип	

Таблица 5. Поддержка генерирования SQL на узле Выборка для простой выборки.

Режим	Пример	Макс. размер	Нач. значен.	DB2 for z/OS	DB2 for OS/400	DB2 for Windows/UNIX	Netezza (смотрите Прим.)	Oracle	SQL Server	Teradata
Включать	Первое	(не задается)		Д	Д	Д	Д	Д	Д	Д
		1-из-п	выключено		Д	Д	Д	Д	Д	Д
	Случ. %	макс	выкл.	выкл.			Д	Д	Д	Д
			вкл.			Д		Д		
		макс	выкл.			Д	Д	Д	Д	Д
			вкл.			Д		Д	Д	

Таблица 5. Поддержка генерирования SQL на узле Выборка для простой выборки (продолжение).

Режим	Пример	Макс. размер	Нач. значен.	DB2 for z/OS	DB2 for OS/400	DB2 for Windows/UNIX	Netezza (смотрите Прим.)	Oracle	SQL Server	Teradata	
Отклонить	Первое	выкл.					Д	Д			
		макс					Д	Д			
	1-из-п	выкл.		Д	Д	Д	Д	Д		Д	
		макс		Д	Д	Д	Д	Д		Д	
	Случ. %	выкл.	выкл.	выкл.			Д	Д	Д		Д
			вкл				Д		Д		
		макс	выкл.				Д	Д	Д		Д
			вкл					Д		Д	

Примечание: Для баз данных Netezza ограничение касается использования ключевых слов в подзапросах, с помощью которых выполняется генерирование SQL. Обходной прием - включить кэширование на узле Выборка.

Таблица 6. Поддержка генерирования SQL на узле Агрегирование.

Хранение	Сумма	Среднее значение	Минимум	Максимум	Стр. отклонение	Медиана	Количество	Дисперсия	Процентиль
Целое число	Д	Д	Д	Д	Д		Д	Д	
Действительное число	Д	Д	Д	Д	Д		Д	Д	
Дата			Д	Д			Д		
Время			Д	Д			Д		
Метка даты/времени			Д	Д			Д		
Строка символов			Д	Д			Д		

Таблица 7. Операции с полями

Узел, поддерживающий генерирование SQL	Примечания
Тип	Поддерживает генерирование SQL, если определен узел Тип и не задана проверка типа ABORT или WARN.
Фильтр	
Произвести от	Поддерживает генерирование SQL, если поддерживается SQL, сгенерированный для вычислительного выражения (смотрите выражения ниже).
Ансамбль	Поддерживает генерирование SQL для непрерывных полей назначения. Для других полей назначения поддерживает генерирование, только если используется метод ансамбля "Побеждает наибольшая доверительная вероятность".
Заполнитель	Поддерживает генерирование SQL, если поддерживается SQL, сгенерированный для вычислительного выражения (смотрите выражения ниже).
Анонимизировать	Поддерживает генерирование SQL для непрерывных полей назначения и частичное генерирование SQL для номинальных и флаговых полей назначения.
Переклассифицировать	

Таблица 7. Операции с полями (продолжение)

Узел, поддерживающий генерирование SQL	Примечания
Категоризация	Поддерживает генерирование SQL, если используется метод разделения на интервалы "Плитки (равное количество)" и выбрана опция "Прочитать со вкладки Значения интервалов разделения, если она доступна".
RFM-анализ (недавность, частота, деньги)	Поддерживает генерирование SQL, если выбрана опция "Прочитать со вкладки Значения интервалов разделения, если она доступна", но нисходящие узлы его не поддерживают.
Подмножества	Поддерживает генерирование SQL, назначая записи разделам.
SetToFlag	
Реструктурировать	

Таблица 8. Графики

Узел, поддерживающий генерирование SQL	Примечания
Панель выбора диаграмм	Генерирование SQL поддерживается для следующих типов диаграмм: область, трехмерная область, столбец, трехмерный столбец, столбец частот, тепловая карта, круг, трехмерный круг, круг частот. Для Гистограмм генерирование SQL поддерживается только для категориальных данных.
Распределение	
Web	
Оценка	

Таблица 9. Моделирование

Узел, поддерживающий генерирование SQL	Примечания
Автоклассификатор	Поддерживает генерирование SQL для непрерывных полей назначения. Для других полей назначения поддерживает генерирование, только если используется метод ансамбля "Побеждает наибольшая доверительная вероятность".
Автоenumerация	

Для некоторых моделей можно сгенерировать SQL для слепка модели, передавая этап скоринга модели обратно в базу данных. Основное назначение этой функции - не повысить производительность, а разрешить потокам, содержащим эти слепки, передавать их полный SQL обратно. Дополнительную информацию смотрите в разделе "Просмотр SQL для слепков моделей" на стр. 43.

Таблица 10. Слепки моделей

Слепок модели, поддерживающий генерирование SQL	Примечания
Дерево C&R	Поддерживает генерирование SQL для опции единичного дерева, но не для опций бустинга, бэггинга или больших наборов данных.
QUEST	
CHAID	
C5.0	
Список решений	
Линейный	Поддерживает генерирование SQL для опции стандартной модели, но не для опций бустинга, бэггинга или больших наборов данных.

Таблица 10. Слелки моделей (продолжение)

Слелок модели, поддерживающий генерирование SQL	Примечания
Нейросеть	Поддерживает генерирование SQL для опции стандартной модели (только многослойный перцептрон), но не для опций бустинга, бэггинга или больших наборов данных.
РСА/фактор	
Логистическое	Поддерживает генерирование SQL для полиномиальной, но не биномиальной процедуры. Для полиномиальной процедуры генерирование не поддерживается, если выбраны доверительные интервалы, а тип поля назначения - не Флаг.
Сгенерированные наборы правил	

Таблица 11. Вывод

Узел, поддерживающий генерирование SQL	Примечания
Таблица	Поддерживает генерирование SQL, если оно поддерживается для выражения выделения (смотрите выражения ниже).
Матрица	Поддерживает генерирование, кроме случаев, когда для опции Поля выбрано "Все числовые".
Анализ	Поддерживает генерирование в зависимости от выбранных опций.
Преобразование	
Статистики	Поддерживает генерирование, если не используется опция Коррелировать.
Отчет	
Задать глобальные параметры	

Таблица 12. Экспорт

Узел, поддерживающий генерирование SQL	Примечания
База данных	
Публикатор	Опубликованный поток будет содержать сгенерированный SQL.

Выражения и операторы CLEM, поддерживающие генерирование SQL

В следующих таблицах показаны математические операции и выражения, поддерживающие генерирование SQL и часто используемые для исследования данных. Операции, отсутствующие в этих таблицах, не поддерживают генерирование SQL в текущем выпуске.

Таблица 13. Операторы

Операция, поддерживающая генерирование SQL	Примечания
+	
-	
/	
*	
><	Используется для конкатенации строк.

Таблица 14. Операторы сравнения

Операция, поддерживающая генерирование SQL	Примечания
=	
/=	Используется, чтобы задать выражение "не равно".
>	
>=	
<	
<=	

Таблица 15. Функции

Операция, поддерживающая генерирование SQL	Примечания
abs	
allbutfirst	
allbutlast	
and	
arccos	
arcsin	
arctan	
arctanh	
cos	
div	
exp	
fracof	
hasstartstring	
hassubstring	
целое	
intof	
isaplhacode	
islowercode	
isnumbercode	
isstartstring	
issubstring	
isuppercode	
last	
length	
locchar	
log	
log10	
lowertoupper	
max	
member	

Таблица 15. Функции (продолжение)

Операция, поддерживающая генерирование SQL	Примечания
min	
negate	
not	
number	
or	
pi	
real	
rem	
round	
sign	
sin	
sqrt	
строка	
strmember	
subscrs	
substring	
substring_between	
uppertolower	
to_string	

Таблица 16. Специальные функции

Операция, поддерживающая генерирование SQL	Примечания
@NULL	
@GLOBAL_AVE	Специальные глобальные функции используются для получения глобальных значений, вычисленных узлом глобальных значений.
@GLOBAL_SUM	
@GLOBAL_MAX	
@GLOBAL_MEAN	
@GLOBAL_MIN	
@GLOBALSDEV	

Таблица 17. Агрегирующие функции

Операция, поддерживающая генерирование SQL	Примечания
Sum	
Mean	
Min	
Max	
Count	

Таблица 17. Агрегирующие функции (продолжение)

Операция, поддерживаемая генерирование SQL	Примечания
SDev	

Использование функций SQL в выражениях CLEM

Функция @SQLFN может использоваться для добавления именованных функций SQL в выражения CLEM только для выполнения в базе данных. Это может быть полезно в особых случаях, где требуется собственный SQL или другие специфичные для поставщика настройки.

Использование этой функции не охватывается стандартным соглашением о поддержке IBM SPSS Modeler, поскольку ее выполнение основано на внешних компонентах базы данных, неподконтрольных IBM Corp., но в особых случаях возможно ее внедрение в качестве дополнительной услуги. За дополнительной информацией, при необходимости, обращайтесь по адресу <http://www.ibm.com/software/analytics/spss/services/>.

Написание запросов SQL

При работе с узлом База данных обратите особое внимание на запросы SQL, приводящие к набору данных с повторяющимися именами столбцов. Эти дублированные имена часто препятствуют оптимизации SQL для любых нисходящих узлов.

IBM SPSS Modeler с помощью вложенных операторов SELECT выполняет обратную передачу SQL для потоков, использующих запросы SQL на узле Источник базы данных. Другими словами, поток вкладывает запрос, заданный в исходном узле Источник базы данных внутрь одного или нескольких операторов SELECT, сгенерированных во время оптимизации нисходящих узлов. Таким образом, если набор результатов запроса содержит повторяющиеся имена столбцов, оператор нельзя сделать вложенным с помощью реляционной СУБД. Трудности с вложением происходят чаще всего во время объединения таблиц, где столбец с одним и тем же именем выбирается в нескольких объединенных таблицах. Например, рассмотрим следующий запрос в узле источника:

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

Такой запрос не позволяет выполнить последующую оптимизацию SQL, поскольку выполнение данного оператора SELECT дало бы в результате два столбца с именем ID.

Когда возникает ситуация с повторяющимися именами столбцов, необходимо более явно писать запросы SQL и определять алиасы столбцов, чтобы разрешить полную оптимизацию SQL. Более явный запрос иллюстрирует следующий оператор:

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

Приложение А. Конфигурирование Oracle на платформах UNIX

Конфигурирование Oracle для оптимизации SQL

При работе сервер IBM SPSS Modeler на платформах UNIX и чтении из базы данных Oracle, следующие советы помогут обеспечить полную оптимизацию сгенерированного SQL в базе данных.

Надлежащая спецификация локали

При работе сервер IBM SPSS Modeler в национальной версии, отличной от поставляемой с драйверами Connect ODBC, необходимо переконфигурировать компьютер для улучшения оптимизации SQL. Драйверы Connect ODBC поставляются только с файлами локали *en_US*. Соответственно, если компьютер сервер IBM SPSS Modeler работает в другой локали или если для оболочки, в которой был запущен сервер IBM SPSS Modeler, не полностью определена локаль, SQL, возможно, не будет полностью оптимизирован в Oracle. Причины этого:

- сервер IBM SPSS Modeler использует файлы локали ODBC, соответствующие локали, в которой он работает, для перевода кодов, возвращенных из базы данных, в текстовые строки. Он тем самым использует эти текстовые строки для определения, с какой базой данных он фактически соединен.
- Если локаль (возвращенная для сервер IBM SPSS Modeler по запросу системы \$LANG) - не *en_US*, IBM SPSS Modeler не сможет преобразовать коды, полученные от драйвера ODBC, в текст. Другими словами, в начале соединения с базой данных в сервер IBM SPSS Modeler возвращается непреобразованный код, а не строка *Oracle*. Это означает, что IBM SPSS Modeler не способен оптимизировать потоки для Oracle.

Чтобы проверить и переустановить спецификации локали:

1. В оболочке UNIX запустите:

```
#locale
```

Будет возвращена информация о локали для оболочки. Например:

```
$ 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. Перейдите в свой каталог Connect ODBC/локаль. (Здесь вы увидите один каталог, *en_US*.)
3. Создайте программную ссылку на этот каталог *en_US*, указав имя настройки локали в оболочке. Пример:

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

Для неанглийской локали, например, *fr_FR.ISO8859-1*, программную ссылку надо создать так:

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

4. После создания ссылки перезапустите сервер IBM SPSS Modeler из той же оболочки. (сервер IBM SPSS Modeler получает информацию о локали от оболочки, из которой он запущен.)

Примечания

При оптимизации компьютера UNIX для обратных переносов SQL в Oracle учтите следующие советы:

- Надо указать полную локаль. В приведенном выше примере ссылка должна иметь вид язык_территория.кодовая_страница. Существующего каталога локали *en_US* недостаточно.
- Для полной оптимизации исследования в базе данных надо задать в оболочке, используемой для запуска сервер IBM SPSS Modeler, и LANG, и LC_ALL. LANG можно определить в оболочке, как и любую другую переменную среды, перед повторным запуском сервер IBM SPSS Modeler. Например, смотрите определение:
`#LANG=en_US.IS08859-15; export LANG`
- Каждый раз при запуске сервер IBM SPSS Modeler необходимо проверить, что информация о локали оболочки полностью задана и что соответствующая программная ссылка существует в каталоге ODBC/локаль.

Приложение В. Конфигурирование сценариев запуска UNIX

Введение

В этом приложении описываются некоторые сценарии, поставляемые с версиями сервера IBM SPSS Modeler для UNIX, и объясняется, как сконфигурировать эти сценарии. Сценарии позволяют:

- Конфигурировать сервер IBM SPSS Modeler для автоматического запуска при перезапуске компьютера сервера.
- Вручную останавливать и перезапускать сервер IBM SPSS Modeler.
- Изменять разрешения для файлов, которые создает сервер IBM SPSS Modeler.
- Конфигурировать сервер IBM SPSS Modeler для работы с драйверами ODBC Connect, поставляемыми с сервером IBM SPSS Modeler. Дополнительную информацию смотрите в разделе “Сервер IBM SPSS Modeler и пакет Data Access Pack” на стр. 56.
- Конфигурировать сервер IBM SPSS Modeler для работы с драйвером IBM SPSS Collaboration and Deployment Services Enterprise View Driver при чтении данных с помощью узла Представление предприятия. Дополнительную информацию смотрите в разделе “Конфигурирование драйвера для узла представления предприятия” на стр. 61.

Сценарии

Сервер IBM SPSS Modeler использует несколько сценариев, в том числе:

- **modelersrv.sh.** Сценарий ручного запуска для сервера IBM SPSS Modeler находится в каталоге, где установлен сервер IBM SPSS Modeler. Он конфигурирует среду для сервера, если процесс демона сервера запущен *вручную*. Запустите его, если нужно вручную запустить или остановить сервер. Если нужно изменить конфигурацию для ручного запуска, отредактируйте этот сценарий.
- **auto.sh.** Это сценарий, конфигурирующий вашу систему для автоматического запуска процесса демона сервера во время загрузки. Запустите его один раз, чтобы сконфигурировать систему для автоматического запуска. Изменять его не нужно. Сценарий находится в каталоге, где установлен сервер IBM SPSS Modeler.
- **rc.modeler.** При запуске *auto.sh* этот сценарий создается в положении, которое зависит от операционной системы вашего сервера. Он конфигурирует среду для сервера, если сервер запущен *автоматически*. Если нужно изменить конфигурацию для автоматического запуска, отредактируйте этот сценарий.

Таблица 18. Положение *rc.modeler* в зависимости от операционной системы

Операционная система	Положение
AIX	<i>/etc/rc.modeler</i>
HP-UX	<i>/sbin/init.d/rc.modeler</i>
Solaris	<i>/etc/init.d/rc.modeler</i>

Автоматический запуск и остановка сервер IBM SPSS Modeler

Сервер IBM SPSS Modeler надо запускать как процесс демона. Программа установки содержит сценарий (*auto.sh*), запуск которого позволяет сконфигурировать систему для автоматической остановки и перезапуска сервера IBM SPSS Modeler.

Чтобы сконфигурировать систему для автоматического запуска и завершения работы

1. Войдите в систему как пользователь *root*.
2. Перейдите в каталог, где установлен сервер IBM SPSS Modeler.
3. Запустите сценарий. В командной строке UNIX введите:


```
./auto.sh
```

Автоматический сценарий запуска *rc.modeler* будет создан в положении, указанном выше в таблице. Операционная система будет использовать *rc.modeler* для запуска процесса демона сервера IBM SPSS Modeler при каждой перезагрузке компьютера сервера. Операционная система будет также использовать *rc.modeler* для остановки демона при каждом завершении работы.

Запуск и остановка сервера IBM SPSS Modeler вручную

Можно вручную запустить и остановить сервер IBM SPSS Modeler, запустив сценарий *modelersrv.sh*.

Чтобы запустить или остановить сервер IBM SPSS Modeler вручную

1. Перейдите в каталог, где установлен сервер IBM SPSS Modeler.
2. Чтобы запустить сервер, введите в командной строке UNIX:

```
./modelersrv.sh start
```
3. Чтобы остановить сервер, введите в командной строке UNIX:

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

Редактирование сценариев

Если используется и ручной, и автоматический запуск, внесите параллельные изменения в оба сценария, *modelersrv.sh* и *rc.modeler*. Если используется только ручной запуск, внесите изменения в *modelersrv.sh*. Если используется только автоматический запуск, внесите изменения в *rc.modeler*.

Чтобы изменить сценарии

1. Остановите сервер IBM SPSS Modeler. (Дополнительную информацию смотрите в разделе “Запуск и остановка сервера IBM SPSS Modeler вручную”.)
2. Найдите нужный сценарий. (Дополнительную информацию смотрите в разделе “Сценарии” на стр. 55.)
3. Откройте сценарий в текстовом редакторе, внесите изменения и сохраните файл.
4. Запустите сервер IBM SPSS Modeler автоматически (путем перезапуска компьютера сервера) или вручную.

Управление разрешениями при создании файла

Сервер IBM SPSS Modeler создает временные файлы с разрешениями на чтение, запись и выполнение для всех пользователей. Эту опцию по умолчанию можно переопределить, изменив значение параметра *UMASK* в одном из сценариев запуска, *modelersrv.sh*, *rc.modeler* или обоих. (Дополнительную информацию смотрите в разделе “Редактирование сценариев” выше.) Для максимального ограничения разрешений рекомендуется использовать для *UMASK* значение 077. Значения с более строгими ограничениями могут вызвать проблемы с доступом к серверу IBM SPSS Modeler.

Сервер IBM SPSS Modeler и пакет Data Access Pack

Если требуется использовать драйверы ODBC с сервером IBM SPSS Modeler, среду ODBC необходимо сконфигурировать с помощью *odbc.sh* при запуске процесса сервера IBM SPSS Modeler. Для этого нужно изменить соответствующий сценарий запуска IBM SPSS Modeler в файле *modelersrv.sh*, *rc.modeler* или обоих файлов. (Дополнительную информацию смотрите в разделе “Редактирование сценариев”.)

Дополнительную информацию смотрите на сайте технической поддержки по адресу <http://www.ibm.com/support>. Для получения дополнительных сведений о создании и настройке разрешений для источников данных ODBC обратитесь к своему администратору базы данных.

Чтобы сконфигурировать ODBC для запуска с сервером IBM SPSS Modeler

1. Остановите хост сервера IBM SPSS Modeler, если он запущен.
2. Скачайте нужный файл архива tar для платформы, на которой установлен сервер IBM SPSS Modeler. Убедитесь, что вы загружаете драйверы для правильной версии сервера IBM SPSS Modeler, установленной у вас. Скопируйте файл в положение, в которое собираетесь установить драйверы ODBC (например, */usr/spss/odbc*).
3. Распакуйте архив.
4. Запустите сценарий *setodbcpath.sh*, извлеченный из архива.
5. Отредактируйте сценарий *odbc.sh*, добавив определение ODBCINI в конец сценария и экспортируйте его, например:

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

В ODBCINI надо указать полный путь к файлу *odbc.ini*, который IBM SPSS Modeler должен прочитать для получения списка заданных вами источников данных ODBC (*odbc.ini* по умолчанию установлен вместе с драйверами).

6. Сохраните файл *odbc.sh*.
7. (Только для 64-битных установок сервера IBM SPSS Modeler, для других установок продолжите со следующего шага) Определите и экспортируйте LD_LIBRARY_PATH_64 в *odbc.sh*:

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

где *путь_библиотеки* совпадает с определением для LD_LIBRARY_PATH, уже содержащемся в сценарии, инициализированном для вашего пути установки (например, */usr/spss/odbc/lib*). Самый простой путь сделать это - скопировать операторы *if* и *export* для LD_LIBRARY_PATH в файл *odbc.sh*, добавить их в конец файла и затем заменить строки "LD_LIBRARY_PATH" во вновь добавленных операторах *if* и *export* на "LD_LIBRARY_PATH_64".

Таким образом, в окончательном виде ваш файл *odbc.sh* в 64-битной установке сервера IBM SPSS Modeler может выглядеть следующим образом:

```
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
```

Не забудьте экспортировать LD_LIBRARY_PATH_64, а также определить его в ветви *if*.

8. Отредактируйте файл *odbc.ini*, который вы определили ранее при помощи \$ODBCINI. Определите имена нужных вам источников данных (они зависят от базы данных, к которой вы обращаетесь).
9. Сохраните файл *odbc.ini*.
10. Сконфигурируйте сервер IBM SPSS Modeler для использования этих драйверов. Для этого отредактируйте файл *modelersrv.sh* и добавьте следующую строку непосредственно под строкой, определяющей SCLEMDNAME:


```
. <путь_odbc.sh>
```

где *путь_odbc.sh* - полный путь к файлу *odbc.sh*, который вы изменили перед началом этой процедуры, например:

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

Примечание: Обратите внимание на синтаксис, между первой точкой и путем к файлу должен быть пробел.

11. Сохраните файл *modelersrv.sh*.

Для проверки соединения

1. Перезапустите сервер IBM SPSS Modeler.
2. Соединитесь с сервером IBM SPSS Modeler с клиента.
3. Добавьте узел Источник базы данных на панель холста на клиенте.
4. Откройте узел и проверьте, что имена источников данных, которые вы ранее определили в файле *odbc.ini* в процедуре конфигурации, видны.

Если вы не видите того, что ожидали увидеть, или получаете ошибки при попытках соединения с источником данных, который вы определили, выполните процедуру устранения неисправностей. Дополнительную информацию смотрите в разделе “Устранение неисправностей конфигурации ODBC”.

Чтобы сконфигурировать ODBC для запуска с модуль времени выполнения IBM SPSS Modeler Solution Publisher

Если вы успешно соединяетесь с базой данных из сервера IBM SPSS Modeler, можно сконфигурировать установку модуль времени выполнения IBM SPSS Modeler Solution Publisher на том же сервере, указав в сценарии запуска модуль времени выполнения IBM SPSS Modeler Solution Publisher ссылку на тот же сценарий *odbc.sh*.

1. Отредактируйте сценарий *modelerrun* в модуль времени выполнения IBM SPSS Modeler Solution Publisher, добавив следующую строку непосредственно перед последней строкой сценария:

```
. <путь_odbc.sh>
```

где *путь_odbc.sh* - полный путь к файлу *odbc.sh*, который вы изменили перед началом этой процедуры, например:

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

Примечание: Обратите внимание на синтаксис, между первой точкой и путем к файлу должен быть пробел.

2. Сохраните файл сценария *modelerrun*.

Чтобы сконфигурировать ODBC для запуска с IBM SPSS Modeler Batch

Для ODBC никакого конфигурирования сценария IBM SPSS Modeler Batch не требуется. Это объясняется тем, что для запуска потоков вы соединяетесь с сервером IBM SPSS Modeler из IBM SPSS Modeler Batch. Убедитесь, что конфигурирование ODBC сервера IBM SPSS Modeler выполнено, и он работает правильно, как описано выше в этом разделе.

Чтобы добавить или изменить имя источника данных

1. Отредактируйте файл *odbc.ini*, добавив в него новое или измененное имя.
2. Проверьте соединение, как описано ранее в этом разделе.

Если соединение с сервером IBM SPSS Modeler работает правильно, новый или измененный источник данных также должен работать правильно с модуль времени выполнения IBM SPSS Modeler Solution Publisher и IBM SPSS Modeler Batch.

Устранение неисправностей конфигурации ODBC

Не выводятся источники данных или выводится случайный текст

Если при открытии узла Источник базы данных список доступных источников данных оказывается пуст или содержит неожиданные записи, это может произойти из-за ошибки в сценарии запуска.

1. Проверьте, что \$ODBCINI определена в *modelersrv.sh*, явным образом в самом сценарии или в сценарии *odbc.sh*, на который ссылается *modelersrv.sh*.
2. В последнем случае убедитесь, что ODBCINI указывает полный путь к файлу *odbc.ini*, с помощью которого вы определили свои источники данных ODBC.
3. Если путь в ODBCINI указан верно, проверьте значение \$ODBCINI, используемое в среде сервера IBM SPSS Modeler, путем эхо-контроля переменной из *modelersrv.sh*. Для этого добавьте в *modelersrv.sh* сразу после определения ODBCINI следующую строку:

```
echo $ODBCINI
```

4. Сохраните и затем выполните файл *modelersrv.sh*. Значение \$ODBCINI, задаваемое в среде сервера IBM SPSS Modeler, будет записано в *stdout* для проверки.
5. Если в *stdout* не возвращено никакого значения и вы определяете \$ODBCINI в сценарии *odbc.sh*, на который сделали ссылку из *modelersrv.sh*, проверьте правильность синтаксиса ссылки. Должно быть:

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

где *путь_odbc.sh* - полный путь к файлу *odbc.sh*, который вы изменили перед началом этой процедуры, например:

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

Примечание: Обратите внимание на синтаксис, между первой точкой и путем к файлу должен быть пробел.

Если в *stdout* после запуска *modelersrv.sh* отображается правильное значение, вам должны быть видны имена источников данных в узле Источник базы данных при перезапуске сервера IBM SPSS Modeler и соединении с ним из клиента.

Клиент IBM SPSS Modeler зависает после нажатия кнопки Соединиться в диалоговом окне Соединения с базой данных

Такое поведение может объясняться неверно заданным путем к библиотекам ODBC в вашем пути библиотек. Путь библиотек задается переменной \$LD_LIBRARY_PATH (и \$LD_LIBRARY_PATH_64 в 64-битных версиях).

Чтобы увидеть значение пути библиотеки в среде демона сервера IBM SPSS Modeler, отобразите значение соответствующей переменной среды из *modelersrv.sh* после строки, добавляющей путь библиотеки ODBC в путь библиотек, и выполните сценарий. Значение пути к библиотеке будет выведено на терминал при следующем выполнении сценария.

Если при настройке среды ODBC сервера IBM SPSS Modeler вы ссылаетесь на *odbc.sh* из *modelersrv.sh*, отобразите значение вашего пути к библиотеке из строки, которая следует за строкой, содержащей ссылку на сценарий *odbc.sh*. Для вывода значения добавьте следующую строку к сценарию, затем сохраните и выполните файл сценария:

```
echo $<переменная_пути_библиотек>
```

где *<переменная_пути_библиотек>* - соответствующая переменная пути к библиотеке для операционной системы вашего сервера.

Возвращенное значение пути библиотек должно содержать путь к подкаталогу *lib* вашей установки ODBC. Если это не так, добавьте этот путь в файл.

При работе с 64-битной версией сервера IBM SPSS Modeler, значение \$LD_LIBRARY_PATH_64 переопределит \$LD_LIBRARY_PATH, если оно было задано. Если эта проблема возникает на одной из этих

64-битных платформ, выведите LD_LIBRARY_PATH_64, а также \$LD_LIBRARY_PATH из файла *modelersrv.sh* и если требуется, задайте \$LD_LIBRARY_PATH_64, указав в определении подкаталог *lib* вашей установки ODBC, и экспортируйте это определение.

Не найдено имя источника данных и не задан драйвер по умолчанию

Если вы видите эту ошибку при нажатии кнопки Соединиться в диалоговом окне Соединения с базой данных, это обычно указывает на то, что ваш *odbc.ini* файл неправильно определен. Убедитесь, что имя источника данных (DSN), заданное в разделе [ODBC Data Sources] в верхней части файла соответствует строке в квадратных скобках ниже в *odbc.ini* для определения DSN. В случае какого-либо несовпадения вы увидите эту ошибку при попытке установить соединение с помощью DSN из IBM SPSS Modeler. Ниже приведен пример *неверной* спецификации:

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

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

Необходимо изменить одну из двух строк, выделенных полужирным шрифтом, чтобы они точно соответствовали друг другу. Это устранит ошибку.

Невозможно загрузить заданный драйвер

Эта ошибка также указывает, что *odbc.ini* файл неправильно определен. Одна из возможностей состоит в том, что параметр Driver в разделе драйвера неверно задан, например:

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

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

1. Проверьте, что совместно используемый объект, заданный в параметре Driver, существует.
2. Исправьте путь к совместно используемому объекту, если он неверен.
3. Если параметр Driver задан в формате:

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

это указывает, что вы не инициализировали связанные с ODBC сценарии. Запустите сценарий *setodbcpath.sh*, установленный вместе с драйверами. Дополнительную информацию смотрите в разделе “Сервер IBM SPSS Modeler и пакет Data Access Pack” на стр. 56. При выполнении этого сценария вы должны вместо строки "ODBCHOME" увидеть появившийся путь к установке ODBC. Это должно решить проблему.

Другой причиной может быть ошибка в библиотеке драйвера. Чтобы подтвердить, что драйвер не удастся загрузить, воспользуйтесь инструментом *ivtestlib*, поставляемым с ODBC. В случае Connect64 используйте инструмент *ddtestlib*. Исправьте ошибку, задав переменную пути библиотек в сценарии запуска.

Например, если драйвер Oracle нельзя загрузить для 32-битной установки, выполните следующие действия:

1. С помощью `ivtestlib` подтвердите, что драйвер не удалось загрузить. Например, в командной строке UNIX введите:

```
sh
cd ODBCDir
. odbcs.sh
./bin/ivtestlib MFor815
```

где вместо ODBCDir указан путь к вашему каталогу установки ODBC.

2. Прочтите сообщение и проверьте, есть ли ошибка. Например, сообщение

Load of MFor815.so failed: ld.so.1: bin/ivtestlib: fatal: libclntsh.so: open failed: No such file or directory
(Загрузка MFor815.so завершилась неудачно: ld.so.1: bin/ivtestlib: неисправимое: libclntsh.so: не удалось открыть: Нет такого файла или каталога)

указывает, что библиотека клиента Oracle, *libclntsh.so*, отсутствует или находится вне пути библиотек (например не в пути LD_LIBRARY_PATH в Solaris).

3. Подтвердите, что библиотека существует. Если это не так, переустановите клиент Oracle. Если библиотека есть, введите следующую последовательность команд в командной строке UNIX:

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

где */bigdisk/oracle/product/8.1.6/lib* заменен путем к *libclntsh.so*, а LD_LIBRARY_PATH - переменная пути библиотек для вашей операционной системы.

Обратите внимание на то, что если вы работаете с 64-битной версией IBM SPSS Modeler в Linux или Solaris, переменная пути библиотек содержит суффикс `_64`. Поэтому первые две строки в предыдущем примере выглядели бы так:

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

4. Прочитайте сообщение, чтобы подтвердить, что драйвер теперь можно загрузить. Например, сообщение

Load of MFor815.so successful, qehandle is 0xFF3A1BE4 (Загрузка MFor815.so успешно завершена, qehandle - 0xFF3A1BE4)

указывает, что можно загрузить библиотеку клиента Oracle.

5. Исправьте путь библиотек в сценарии запуска IBM SPSS Modeler.
6. Перезапустите сервер IBM SPSS Modeler, используя отредактированный вами сценарий запуска (*modelersrv.sh* или *rc.modeler*).

Пути библиотек

Имя переменной пути библиотеки зависит от вашей операционной системы. При конфигурировании системы и устранении в ней неисправностей сделать правильные замены поможет следующая таблица.

Таблица 19. Путь библиотеки в зависимости от операционной системы

Операционная система	Имя переменной пути библиотеки
AIX	LIBPATH
HP-UX	SHLIB_PATH
Solaris или Linux 32-битная	LD_LIBRARY_PATH
Solaris или Linux 64-битная	LD_LIBRARY_PATH_64

Конфигурирование драйвера для узла представления предприятия

Узел представления предприятия позволяет создавать и поддерживать соединение между сеансом IBM SPSS Modeler и IBM SPSS Collaboration and Deployment Services Enterprise View в совместно используемом репозитории IBM SPSS Collaboration and Deployment Services Repository.

Чтобы использовать узел представления предприятия с сервером IBM SPSS Modeler необходимо установить и сконфигурировать IBM SPSS Collaboration and Deployment Services Enterprise View Driver. Для получения дополнительной информации по установке драйвера обратитесь к своему администратору. После установки необходимо добавить в сценарий запуска (в файл *modelersrv.sh*, *rc.modeler* или в оба файла) ссылку на *pev.sh* в формате:

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

Приложение С. Конфигурирование и запуск сервер SPSS Modeler в качестве процесса обычного пользователя (не root) в UNIX

Введение

Эти инструкции содержат информацию о том, как сервер IBM SPSS Modeler выполняется в качестве процесса обычного пользователя (не root) в системах UNIX.

Выполнение в качестве процесса пользователя root. Если сервер IBM SPSS Modeler установлен с опциями по умолчанию, процесс демона сервера будет выполняться как процесс пользователя root. Выполнение процесса как root позволяет IBM SPSS Modeler надежно аутентифицировать каждого пользователя при входе в систему и начинать сеанс каждого пользователя под соответствующей учетной записью UNIX. Это обеспечивает пользователям доступ только к их собственным файлам и каталогам.

Выполнение в качестве процесса обычного пользователя (не root). Выполнение сервера IBM SPSS Modeler как процесса обычного пользователя (не root) означает, что фактический действующий ID пользователя процесса демона на сервере задаются выбранной вами учетной записью. Все пользовательские сеансы, которые запускает сервер SPSS Modeler, будут использовать одну и ту же учетную запись UNIX, что означает, что любые данные файлов, прочитанные или записанные SPSS Modeler, будут совместно использоваться пользователями SPSS Modeler. Это не повлияет на доступ к данным базы данных, поскольку пользователям нужно пройти аутентификацию независимо от того, какой из источников данных базы данных они используют. Без привилегии root IBM SPSS Modeler может работать одним из двух способов:

- **Без базы данных секретных паролей.** Этот метод SPSS Modeler использует при работе с существующей базой данных паролей UNIX, NIS или сервером LDAP, которые обычно используются для аутентификации пользователя в системе UNIX. Дополнительную информацию смотрите в разделе “Конфигурирование в качестве обычного пользователя (не root) без помощи базы данных секретных паролей”.
- **С базой данных секретных паролей.** Этот метод применяется SPSS Modeler для аутентификации пользователей в базе данных секретных паролей, отличной от базы данных паролей UNIX, NIS или LDAP, которые обычно используются для аутентификации пользователя в UNIX. Дополнительную информацию смотрите в разделе “Конфигурирование в качестве рядового пользователя (не root) с помощью базы данных секретных паролей” на стр. 64.

Конфигурирование в качестве обычного пользователя (не root) без помощи базы данных секретных паролей

Чтобы сконфигурировать сервер IBM SPSS Modeler для запуска с учетной записью без полномочий root, без использования базы данных секретных паролей, выполните следующие действия:

1. Откройте в редакторе файл сервера SPSS Modeler *options.cfg*.
2. Задайте для опции `start_process_as_login_user` значение Y.
3. Сохраните и закройте файл *options.cfg*.

Заметим, что запуск в качестве обычного пользователя (не root), скорее всего, потребует изменения конфигурации. Дополнительную информацию смотрите в разделе “Устранение неисправностей при аутентификации пользователей” на стр. 65.

Предостережение: Не включайте опцию `start_process_as_login_user`, если вы затем запускаете сервер IBM SPSS Modeler как пользователь *root*. В таком случае для всех пользователей, соединенных с сервером, их серверные процессы будут выполняться как *root*, что связано с риском для безопасности.

Конфигурирование в качестве рядового пользователя (не root) с помощью базы данных секретных паролей

Если вы решили аутентифицировать пользователей посредством базы данных секретных паролей, все пользовательские сеансы будут запущены для одной и той же учетной записи пользователя без полномочий root.

Чтобы сконфигурировать сервер IBM SPSS Modeler для запуска с учетной записи без полномочий root, выполните следующие действия:

1. Создайте группу, содержащую всех ваших пользователей. Имя этой группы может быть любым, в данном примере используется имя *modelerusers*.
2. Создайте учетную запись пользователя, под которой будет работать сервер IBM SPSS Modeler. Эта учетная запись предназначена для монопольного использования процессом демона сервера IBM SPSS Modeler. В этом примере она будет называться *modelerserv*.

При создании учетной записи учтите, что:

- Первичная группа должна быть ранее созданной группой *<modelerusers>*.
- Домашним каталогом может быть каталог установки IBM SPSS Modeler или любой другой удобный каталог по умолчанию (выберите каталог, отличный от каталога установки, если учетная запись должна сохраняться при обновлениях).

3. Затем сконфигурируйте сценарии запуска, чтобы запускать сервер IBM SPSS Modeler при помощи вновь созданной учетной записи. Найдите подходящий сценарий запуска и откройте его в текстовом редакторе. Дополнительную информацию смотрите в разделе “Сценарии” на стр. 55.

- a. Измените значение параметра *umask*, чтобы разрешить, как минимум, групповой доступ для чтения к созданным файлам:

```
umask 027
```

4. Отредактируйте файл опций сервера *config/options.cfg*, чтобы задать аутентификацию в базе данных секретных паролей, добавив строку:

```
authentication_methods, "pasw_modeler"
```

5. Задайте для опции *start_process_as_login_user* значение Y.

6. Затем создайте базу данных секретных паролей, сохранив ее в файле *config/passwords.cfg*. В файле паролей заданы комбинации имени пользователя и пароля, которым разрешен вход в систему IBM SPSS Modeler. *Примечание:* Они секретны для IBM SPSS Modeler и не связаны с именами пользователей и паролями для входа в систему UNIX. Можно для удобства использовать одни и те же имена пользователей, но нельзя использовать те же пароли.

Чтобы создать файл паролей, нужна утилита паролей, *pwutil*, которая находится в подкаталоге *bin* каталога, где установлен сервер IBM SPSS Modeler. Синтаксис этой программы:

```
pwutil [ имя_пользователя [ пароль ] ]
```

Программа получает имя пользователя и незашифрованный пароль и записывает имя пользователя и зашифрованный пароль в стандартный вывод в формате, подходящем для файла паролей. Например, чтобы задать пользователя *modeler* с паролем "data mining", нужно ввести:

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

В большинстве случаев, если все пользователи входят в систему с одним и тем же именем и паролем, достаточно задать одно имя пользователя. Однако с помощью оператора *>>* можно создать дополнительных пользователей, добавив их в этот файл, например:

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

Примечание: Если используется одинарный символ *>*, содержимое *passwords.cfg* будет каждый раз перезаписываться с заменой ранее заданных пользователей. Не забудьте, что при этом все пользователи совместно используют одну и ту же учетную запись пользователя UNIX.

Примечание: Если добавить новых пользователей в базу данных секретных паролей, пока сервер SPSS Modeler работает, надо перезапустить сервер SPSS Modeler, чтобы вновь заданные пользователи могли

быть распознаны системой. Пока вы этого не сделаете, вход в систему пользователей, добавленных при помощи `rwutil` с момента, когда сервер SPSS Modeler был перезапущен последний раз, будет невозможен.

7. Рекурсивно измените владельца каталога установки IBM SPSS Modeler и его содержимого на пользователя `<modelerserv>` и группу `<modelerusers>`, где указаны имена, созданные вами ранее. Например:

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

8. Для пользователей IBM SPSS Modeler имеет смысл создать подкаталоги в каталоге данных, чтобы они могли сохранять рабочие данные, не мешая друг другу. Владелец этих каталогов должна быть группа `<modelerusers>` с полномочиями группового чтения, записи и поиска. Например, чтобы создать рабочий каталог для пользователя `bob`:

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

Кроме того, можно задать бит, задающий групповой ID таким образом, чтобы владельцем файлов данных, копируемых в этот каталог, автоматически становилась группа `<modelerusers>`:

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

Запуск сервера SPSS Modeler с учетной записью обычного пользователя (не root)

Чтобы запустить сервер SPSS Modeler от имени обычного пользователя (не root), выполните следующие действия:

1. Войдите в систему с учетной записью пользователя без полномочий root, созданной ранее.
2. Если вы работаете с включенной опцией файла конфигурации `start_process_as_login_user`, можно запускать, останавливать и проверять состояние сервера SPSS Modeler. Дополнительную информацию смотрите в разделе “Для запуска, остановки и проверки состояния в UNIX” на стр. 21.

Конечные пользователи соединяются с сервером SPSS Modeler входя в систему из клиентских программ. Необходимо дать пользователям информацию, которая требуется для соединения, включая IP-адрес или имя хоста компьютера сервера.

Запуск сервера IBM SPSS Modeler с учетной записью пользователя без полномочий root в AIX

При запуске сервера IBM SPSS Modeler от имени пользователя без полномочий root в AIX существует возможность, что на консоль будет выведена ошибка `lsattr not found`. Это связано с тем, что `lsattr` находится в каталоге `/usr/sbin`, но когда на сервер IBM SPSS Modeler входит новый пользователь (для запуска без полномочий root), путь к каталогу сбрасывается и содержит уже только `/usr/bin` и `/bin`.

Если возникла эта ошибка, необходимо создать следующую ссылку в каталоге `/usr/bin`:

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

После создания ссылки перезапустите сервер IBM SPSS Modeler.

Устранение неисправностей при аутентификации пользователей

В зависимости от того, как операционная система сконфигурирована для выполнения аутентификации, возможны ошибки при попытках войти на сервер SPSS Modeler при работе в конфигурации, отличной от root. Например, это возможно, если операционная система сконфигурирована (с помощью `/etc/nsswitch.conf` или подобного файла) для проверки локального теневого файла паролей вместо NIS или LDAP. Это происходит потому, что сервер SPSS Modeler требует доступа для чтения к файлам, служащим для

выполнения аутентификации, включая файл */etc/shadow* или эквивалентный ему (например, */etc/security/passwd* в AIX), где хранится защищенная информация об учетной записи пользователя. Однако разрешения для файла операционной системы обычно заданы так, чтобы каталог */etc/shadow* был доступен только пользователю root. При этих обстоятельствах процесс другого пользователя (не root) не может читать */etc/shadow* для проверки паролей пользователей, что приводит к ошибке аутентификации.

Есть несколько вариантов решения этой проблемы:

- Попросите администратора системы сконфигурировать операционную систему для использования NIS или LDAP для аутентификации.
- Измените разрешения для защищенных файлов, например, предоставив доступ для чтения файлу */etc/shadow*, чтобы локальная учетная запись пользователя, с которой запускался сервер SPSS Modeler, получила доступ к этому файлу. Хотя это обходное решение может показаться неподходящим в производственных средах, его можно временно применить к среде тестирования, чтобы проверить, связана ли ошибка авторизации с конфигурацией операционной системы.
- Задайте список управления доступом (access control list, ACL) для файла */etc/shadow* (смотрите пример ниже в этом разделе).
- Запустите сервер SPSS Modeler от имени пользователя root, чтобы процессы сервера могли читать файл */etc/shadow*. *Предостережение:* В этом случае во избежание описанных выше проблем с защитой проверьте, что в файле *options.cfg* для сервера SPSS Modeler есть опция `start_process_as_login_user`, N.

Пример списка управления доступом (access control list, ACL)

В операционных системах Solaris создать список управления доступом можно с помощью команды `setfacl` (узнайте у своего администратора, доступна ли эта команда в вашей системе). В следующем примере показано, как создать ACL для пользователя *modserv*:

```
# vi /tmp/shadow.ac1
user::r--
user:modserv:r--
group:--
mask:r--
other:--

# setfacl -f /tmp/shadow.ac1 /etc/shadow

# getfacl /etc/shadow

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

Приложение D. Конфигурирование и запуск сервера SPSS Modeler при помощи файла секретных паролей в Windows

Введение

Эти инструкции содержат информацию о том, как сервер IBM SPSS Modeler выполняется с использованием секретного пароля в системах Windows. Этот метод применяется IBM SPSS Modeler для аутентификации пользователей с помощью базы данных секретных паролей, а не с помощью системной аутентификации Windows.

Конфигурирование с использованием базы данных секретных паролей

Если вы решили аутентифицировать пользователей посредством базы данных секретных паролей, все пользовательские сеансы будут запущены для одной и той же учетной записи пользователя.

Чтобы сконфигурировать сервер IBM SPSS Modeler таким образом, выполните следующие действия:

1. Создайте учетную запись пользователя, под которой будет работать сервер IBM SPSS Modeler. Эта учетная запись предназначена для монопольного использования процессом демона сервера IBM SPSS Modeler. В этом примере она будет называться *modelerserv*.
2. Отредактируйте файл опций сервера *config/options.cfg*, чтобы задать для опции `start_process_as_login_user` значение `Y` и указать аутентификацию с использованием базы данных секретных паролей, добавив строку:
`authentication_methods, "pasw_modeler"`
3. Затем создайте базу данных секретных паролей, сохранив ее в файле *config/passwords.cfg*. В файле паролей заданы комбинации имени пользователя и пароля, которым разрешен вход в систему IBM SPSS Modeler. *Примечание:* Они секретны для IBM SPSS Modeler и не связаны с именами пользователей и паролями для входа в систему Windows. Можно для удобства использовать одни и те же имена пользователей, но нельзя использовать те же пароли.

Чтобы создать файл паролей, нужна утилита паролей, *pwutil*, которая находится в подкаталоге *bin* каталога, где установлен сервер IBM SPSS Modeler. Синтаксис этой программы:

```
pwutil [ имя_пользователя [ пароль ] ]
```

Программа получает имя пользователя и незашифрованный пароль и записывает имя пользователя и зашифрованный пароль в стандартный вывод в формате, подходящем для файла паролей. Например, чтобы задать пользователя *modeler* с паролем "data mining", нужно ввести:

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

В большинстве случаев, если все пользователи входят в систему с одним и тем же именем и паролем, достаточно задать одно имя пользователя. Однако с помощью оператора `>>` можно создать дополнительных пользователей, добавив их в этот файл, например:

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

Примечание: Если используется одинарный символ `>`, содержимое *passwords.cfg* будет каждый раз перезаписываться с заменой ранее заданных пользователей. Не забудьте, что при этом все пользователи совместно используют одну и ту же учетную запись пользователя UNIX.

Примечание: Если добавить новых пользователей в базу данных секретных паролей, пока сервер SPSS Modeler работает, надо перезапустить сервер SPSS Modeler, чтобы вновь заданные пользователи могли быть распознаны системой. Пока вы этого не сделаете, вход в систему пользователей, добавленных при помощи *pwutil* с момента, когда сервер SPSS Modeler был перезапущен последний раз, будет невозможен.

4. Дайте пользователю, созданному на шаге 1, полные права управления файлом опций сервера *config/options.cfg* и каталогом *%ALLUSERSPROFILE%\IBM\SPSS*.
5. На панели системных служб остановите службу *IBM SPSS Modeler Server*, измените значение в столбце *Вход от имени* с *Локальная система* на учетную запись пользователя, созданного на шаге 1, и перезапустите эту службу.

Приложение Е. Балансировка нагрузок при помощи кластеров серверов

Для управления службами в сети с IBM SPSS Collaboration and Deployment Services может использоваться подключаемый модуль, называемый координатором процессов. Координатор процессов содержит возможности управления сервером, разработанные для оптимизации взаимодействия клиента с сервером и обработки.

Службы под управлением, такие как сервер IBM SPSS Statistics или сервер IBM SPSS Modeler, регистрируются на координаторе процессов после запуска и периодически посылают ему свежие сообщения о состоянии. Службы могут также хранить необходимые файлы конфигурации в IBM SPSS Collaboration and Deployment Services Repository и получать их при инициализации.

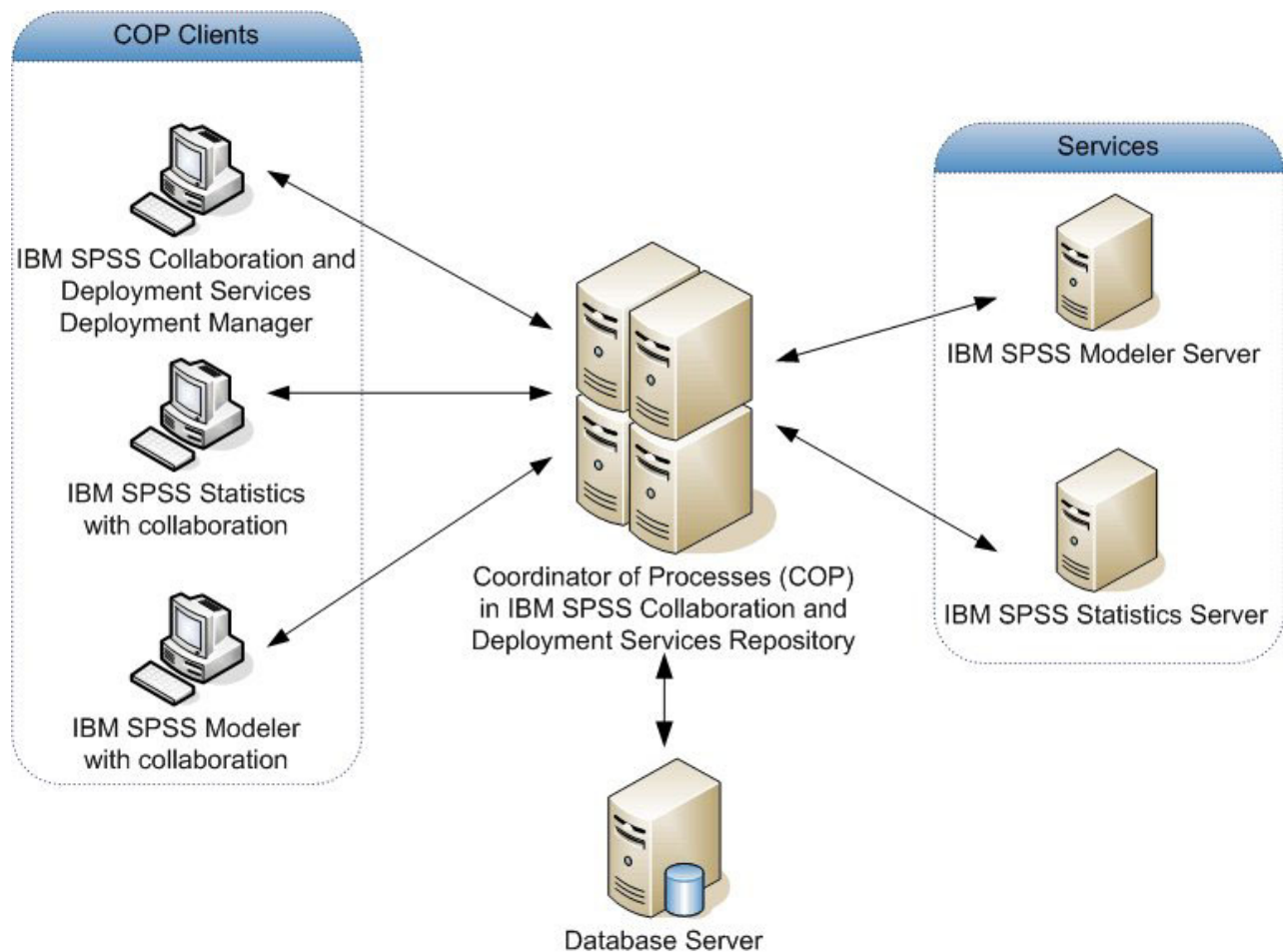


Рисунок 5. Архитектура координатора процессов

Выполнение потоков IBM SPSS Modeler на сервере может повысить производительность. В некоторых случаях можно выбирать только между использованием одного или двух серверов. В других случаях вам может быть предложен больший выбор серверов, поскольку между серверами могут быть существенные различия по таким характеристикам, как владелец, права доступа, данные сервера, тип сервера (тестовый или производственный) и так далее. Кроме того, если в вашей сети есть координатор процессов, вам может быть предложен кластер серверов.

Кластер серверов - это группа серверов, которые являются взаимозаменяемыми с точки зрения конфигурации и ресурсов. Координатор процессов определяет, какой сервер лучше всего подходит для ответа на требование обработки, с помощью алгоритма, который распределит нагрузку с учетом нескольких критериев, включая вес сервера, пользовательские приоритеты и текущую нагрузку обработки.

Дополнительную информацию смотрите в руководстве *Coordinator of Processes Service Developer's Guide* (Руководство разработчика службы координатора процессов), предоставляемом по требованию IBM Corp.

При соединении с сервером или кластером серверов в IBM SPSS Modeler вы можете ввести сервер вручную или найти сервер или кластер с помощью координатора процессов. Дополнительную информацию смотрите в разделе “Соединение с сервер IBM SPSS Modeler” на стр. 15.

Уведомления

Эта информация относится к продуктам и сервису, предлагаемым по всему миру.

IBM может не предоставлять в других странах продукты, услуги и аппаратные средства, описанные в данном документе. За информацией о продуктах и услугах, предоставляемых в вашей стране, обращайтесь к местному представителю IBM. Ссылки на продукты, программы или услуги IBM не означают и не предполагают, что можно использовать только указанные продукты, программы или услуги IBM. Разрешается использовать любые функционально эквивалентные продукты, программы или услуги, если при этом не нарушаются права IBM на интеллектуальную собственность. Однако ответственность за оценку и проверку работы любого продукта, программы или сервиса, не произведенного корпорацией IBM, лежит на пользователе.

IBM может располагать патентами или рассматриваемыми заявками на патенты, относящимися к предмету данного документа. Предъявление данного документа не предоставляет какую-либо лицензию на эти патенты. Вы можете послать письменный запрос о лицензии по адресу:

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

По поводу лицензий, связанных с использованием наборов двухбайтных символов (DBCS), обращайтесь в отдел интеллектуальной собственности IBM в вашей стране или направьте запрос в письменной форме по адресу:

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

Следующий абзац не применяется в Великобритании или в любой другой стране, где подобные заявления противоречат местным законам: INTERNATIONAL BUSINESS MACHINES CORPORATION ПРЕДСТАВЛЯЕТ ДАННУЮ ПУБЛИКАЦИЮ "КАК ЕСТЬ", БЕЗ КАКИХ-ЛИБО ГАРАНТИЙ, КАК ЯВНЫХ, ТАК И ПОДРАЗУМЕВАЕМЫХ, ВКЛЮЧАЯ, НО НЕ ОГРАНИЧИВАЯСЬ ТАКОВЫМИ, ПРЕДПОЛАГАЕМЫЕ ГАРАНТИИ СОБЛЮДЕНИЯ ЧЬИХ-ЛИБО АВТОРСКИХ ПРАВ, ВОЗМОЖНОСТИ КОММЕРЧЕСКОГО ИСПОЛЬЗОВАНИЯ ИЛИ ПРИГОДНОСТИ ДЛЯ КАКИХ-ЛИБО ЦЕЛЕЙ И СООТВЕТСТВИЯ ОПРЕДЕЛЕННОЙ ЦЕЛИ. В некоторых штатах при определенных соглашениях не допускается отказ от выраженных или подразумеваемых гарантий, поэтому данное заявление может к вам не относиться.

Эта информация может содержать технические неточности и типографские ошибки. В представленную здесь информацию периодически вносятся изменения; эти изменения будут включаться в новые издания данной публикации. Фирма IBM может в любое время без уведомления вносить изменения и усовершенствования в продукты и программы, описанные в этой публикации.

Любые приводимые здесь ссылки на web-сайты, не относящиеся к компании IBM, даются исключительно для удобства и ни в коей мере не служат целям поддержки или рекламы этих web-сайтов. Материалы этих Web-сайтов не являются частью данного продукта IBM, и вы можете использовать их только на собственную ответственность.

Любую предоставленную вами информацию IBM может использовать или распространять любым способом, какой сочтет нужным, не беря на себя никаких обязательств по отношению к вам.

Если обладателю лицензии на данную программу понадобятся сведения о возможности: (i) обмена данными между независимо разработанными программами и другими программами (включая данную) и (ii) совместного использования таких данных, он может обратиться по адресу:

IBM Software Group
ATTN: Licensing
200 W. Madison St.
Chicago, IL; 60606
U.S.A.

Такая информация может быть доступна при соответствующих условиях и соглашениях, включая в некоторых случаях взимание платы.

Описанную в данном документе лицензионную программу и все прилагаемые к ней лицензированные материалы IBM предоставляет на основе положений Соглашения между IBM и Заказчиком, Международного Соглашения о Лицензиях на Программы IBM или любого эквивалентного соглашения между IBM и заказчиком.

Любые данные о выполнении, содержащиеся здесь, были определены в контролируемой среде. Поэтому результаты, полученные в других операционных средах, могут существенно отличаться. Некоторые измерения могли быть сделаны на системах в стадии разработки, и поэтому нет гарантии, что соответствующие показатели останутся теми же на общедоступных системах. Более того, некоторые показатели могли быть оценены путем экстраполяции. Реальные результаты могут отличаться. Пользователи этого документа должны проверить приводимые данные в их конкретной среде.

Информация о продуктах, не принадлежащих компании IBM, была получена от поставщиков этих продуктов, из их опубликованных сообщений или других общедоступных источников. Компания IBM не тестировала эти продукты и не может подтвердить правильность их работы, совместимость и другие утверждения, касающиеся продуктов, не принадлежащих компании IBM. Вопросы о возможностях этих продуктов следует направлять их поставщикам.

Все заявления, касающиеся будущих направлений деятельности или намерений корпорации IBM, подвержены изменению или отмене без предупреждения и являются не более чем выражением целей или намерений.

Эти сведения содержат примеры данных и отчетов, используемых в повседневных деловых операциях. Чтобы проиллюстрировать их настолько полно, насколько это возможно, данные примеры включают имена индивидуумов, названия компаний, брендов и продуктов. Все эти имена и названия являются вымышленными, и любое совпадение с названиями и адресами, используемыми реально действующими компаниями, является чисто случайными.

При просмотре данного электронного информационного документа фотографии и цветные иллюстрации могут не показываться.

Товарные знаки

IBM, логотип IBM, и ibm.com являются товарными знаками или зарегистрированными товарными знаками компании International Business Machines Corp., зарегистрированными во многих странах мира. Прочие наименования продуктов и услуг могут быть товарными знаками, принадлежащими IBM или другим компаниям. Текущий список товарных знаков IBM можно найти в Интернете “Copyright and trademark information” по адресу: www.ibm.com/legal/copytrade.shtml.

Intel, логотип Intel, Intel Inside, логотип Intel Inside, Intel Centrino, логотип Intel Centrino, Celeron, Intel Xeon, Intel SpeedStep, Itanium и Pentium являются товарными знаками или зарегистрированными товарными знаками компании Intel или ее дочерних компаний в Соединенных Штатах и других странах.

Linux является зарегистрированным товарным знаком Linus Torvalds в Соединенных Штатах и других странах.

Microsoft, Windows, Windows NT и логотип Windows являются товарными знаками корпорации Microsoft в Соединенных Штатах и других странах.

UNIX является зарегистрированным товарным знаком The Open Group в Соединенных Штатах и других странах.

Java и все основанные на Java товарные знаки и логотипы - товарные знаки или зарегистрированные товарные знаки Oracle и/или его филиалов.

Другие названия продуктов и услуг могут являться товарными знаками IBM или других компаний.

Индекс

Спец. символы

-password
сервер IBM SPSS Modeler 15

Числа

64-битные операционные системы 9

A

allow_modelling_memory_override
 файл options.cfg 25
auto.sh (UNIX)
 положение 55

C

cache_compression
 файл options.cfg 25
chemsrv.sh (UNIX)
 положение 55
COP 18
 балансировка нагрузок 69
 кластеры серверов 69
cop_enabled
 файл options.cfg 27
cop_host
 файл options.cfg 27
cop_password
 файл options.cfg 27
cop_port_number
 файл options.cfg 27
cop_service_description
 файл options.cfg 27
cop_service_host
 файл options.cfg 27
cop_service_name
 файл options.cfg 27
cop_service_weight
 файл options.cfg 27
cop_update_interval
 файл options.cfg 27
cop_user_name
 файл options.cfg 27

D

Data Access Pack
 ODBC, конфигурирование в UNIX 56
 и пути библиотек UNIX 61
 конфигурирование UNIX для 56
 устранение неисправностей ODBC в
 UNIX 58
data_file_path
 файл options.cfg 25
data_files_restricted
 файл options.cfg 25
DB2
 оптимизация SQL 39, 40

I

IBM SPSS Modeler 1
 документация 3
IBM SPSS Modeler Administration
 Console 23
IBM SPSS Modeler Server
 аутентификация пользователей 19
 информация для конечных
 пользователей 18
 каталог temp 25
 координатор конфигураций
 процессов 27
 номер порта 24
 опции администрирования 22
 учетные записи пользователей 19
ID пользователя
 сервер IBM SPSS Modeler 15
io_buffer_size
 файл options.cfg 25

M

max_file_size
 файл options.cfg 25
max_login_attempts
 файл options.cfg 24
max_parallelism
 файл options.cfg 25
max_sessions
 файл options.cfg 24
max_sql_string_length
 файл options.cfg 27
memory_usage
 файл options.cfg 25
Microsoft SQL Server
 оптимизация SQL 39, 40
modelling_memory_limit_percentage
 файл options.cfg 25

O

ODBC
 конфигурирование в UNIX 56
Oracle
 оптимизация SQL 39, 40, 53

P

port_number
 файл options.cfg 24
program_file_path
 файл options.cfg 25
program_files_restricted
 файл options.cfg 25

R

RAM 10

rc.modeler (UNIX)
 положение 55

S

SQL
 запросы 51
 оптимизация Oracle 53
 повторяющиеся имена столбцов 51
 предварительный просмотр
 сгенерированного 43
 просмотр для слепков моделей 43
SQL Server
 оптимизация SQL 39, 40
sql_generation_enabled
 файл options.cfg 27
SSL 30
 обеспечение защищенной связи 30
 обзор 30
ssl_certificate_file
 файл options.cfg 27
ssl_enabled
 файл options.cfg 27
ssl_private_key_file
 файл options.cfg 27
ssl_private_key_password
 файл options.cfg 27
stream_rewriting_enabled
 файл options.cfg 25

T

temp_directory
 файл options.cfg 25

U

UNIX
 аутентификация пользователей 19
 единая регистрация 17
 конфигурирование разрешений для
 файла 56
 пути библиотек 61

A

автоматический запуск сервера
 конфигурирование в UNIX 55
администрирование
 сервера IBM SPSS Modeler 22
архитектура
 компоненты 7
аутентификация 19
аутентификация пользователей 19

Б

баз данных
 доступ 11

В

- временные файлы 10
- разрешения для (IBM SPSS Modeler Server) 19
- вход в сервер IBM SPSS Modeler 15
- выполнение нескольких потоков 25
- выражения CLEM
 - генерирование SQL 48

Г

- генерирование SQL 39, 40, 44
 - включение 42
 - включение для сервера IBM SPSS Modeler 27
 - выражения CLEM 43, 48
 - запись в журнал 43
 - перезапись потока 43
 - предварительный просмотр 43
 - просмотр для слепков моделей 43
 - советы 43

Д

- добавление соединений с сервером IBM SPSS Modeler 17, 18
- документация 3
- доступ администратора для IBM SPSS Modeler Server 23
- доступ к данным 11

Е

- единая регистрация 15

Ж

- жесткие диски 10

З

- запуск IBM SPSS Modeler Server
 - в UNIX 21
 - в Windows 21
- защита
 - SSL 30
 - конфигурирование создания файла в UNIX 56
 - создание файлов 19
- зомби-процессы, сервер IBM SPSS Modeler 22

И

- имена файлов
 - UNIX 12
 - Windows 12
- имена файлов UNC 12
- имя домена (Windows) сервер IBM SPSS Modeler 15
- имя хоста
 - сервер IBM SPSS Modeler 15, 17
- исследование в базе данных 37

- источники данных ODBC
 - ODBC и сценарии UNIX 56
 - и UNIX 56

К

- каталог temp
 - для IBM SPSS Modeler Server 25
- консоль администрирования IBM SPSS Modeler
 - доступ администратора 23
- конфигурация COP
 - для сервера IBM SPSS Modeler 27
- координатор конфигураций процессов для сервера IBM SPSS Modeler 27
- координатор процессов 18
 - балансировка нагрузок 69
 - кластеры серверов 69
- кэширование базы данных
 - генерирование SQL 43
 - управление с помощью файла options.cfg 29
- кэширование в базе данных 29
- кэширование узлов
 - генерирование SQL 43
 - запись в базу данных 43
- кэширование, в базе данных 29

М

- моделирование
 - управление памятью 25

Н

- номер порта сервер IBM SPSS Modeler 15, 17, 24

О

- обратные переносы 39, 40, 42, 44
 - выражения CLEM 48
- обратный перенос SQL. Смотрите также генерирование SQL 39
- ограничения ядра UNIX 21
- ограничения ядра в UNIX 21
- операции
 - генерирование SQL 48
- операционные системы
 - 64-битные 9
- оптимизация
 - генерирование SQL 39, 40, 42
- опции конфигурации COP 27
 - автоматическое генерирование SQL 27
 - длина строки SQL 27
 - доступ к файлам данных 25
- каталог temp 25
- координатор процессов 27
- номер порта 24
- обзор 24
- параллельная обработка
 - перезапись потока 25
 - попытки входа в систему 24

- опции конфигурации (*продолжение*)
 - производительность и оптимизация 25
 - сервера IBM SPSS Modeler 22
 - соединения и сеансы 24
 - управление памятью 25
 - шифрование данных SSL 27
- опция cache_connection 29
- остановка IBM SPSS Modeler Server
 - в UNIX 21
 - в Windows 21
- ошибка при выполнении потока 25

П

- память 10
- параллельная обработка
 - управление 25
- параметры брандмауэра settings
 - файл options.cfg 25
- параметры порта сервера
 - файл options.cfg 25
- параметры портов
 - файл options.cfg 25
- перезапись потока 43
- поиск соединений в COP 18
- предварительный просмотр генерирование SQL 43
- примеры
 - обзор 4
 - Руководство по прикладным программам 3
- примеры прикладных программ 3
- производительность сервера IBM SPSS Modeler 35
- пространство на диске 10
 - вычисление 10
- протокол SSL 30
- процессоры 9
 - несколько 25
- процессы, неотвечающие 22
- пути 12

Р

- разрешения для файла
 - конфигурирование в UNIX 56
- режим доступа к файлам на сервере IBM SPSS Modeler 19
- результаты
 - округление 19
 - порядок записей 19
 - различия между клиентом и сервером 19
- рекомендации по аппаратным средствам для IBM SPSS Modeler Server 9

С

- сервер
 - добавление соединений 17
 - единая регистрация 16, 17
 - подключиться к 15
 - поиск серверов в COP 18
- сервер IBM SPSS Analytic
 - опции конфигурации 24

- сервер IBM SPSS Modeler 1
 - password 15
 - ID пользователя 15
 - администрирование 23
 - доступ администратора 23
 - единая регистрация 16, 17
 - имя домена (Windows) 15
 - имя хоста 15, 17
 - использование мониторинга 29
 - конфигурация COP 27
 - неотвечающие процессы 22
 - номер порта 15, 17
 - опции конфигурации 24
 - процессы сервера 29
 - различные результаты на клиенте и на сервере 19
 - создание файлов 19
- серверы баз данных 37
- сжатие кэша 25
- слепки моделей
 - просмотр SQL для 43
- соединения
 - в сервер IBM SPSS Modeler 15, 17, 18
 - кластер сервера 18
- соединения с базами данных
 - закрытие 29
- сообщения
 - вывод сгенерированного SQL 43
- состояние
 - сервера IBM SPSS Modeler в UNIX 21
 - сервера IBM SPSS Modeler в Windows 21
- сценарии UNIX
 - auto.sh 55
 - modelersrv.sh 55
 - rc.modeler 55
 - редактирование 56

Ш

- шифрование
 - FIPS 28
 - SSL 30
- шифрование FIPS 28
- шифрование данных SSL
 - включение для сервера IBM SPSS Modeler 27

Т

- технология доступа к данным IBM SPSS Statistics 11

У

- узлы
 - поддерживающие генерирование SQL 44
- управление памятью
 - опции администрирования 25
- учетные записи пользователей
 - сервер IBM SPSS Modeler 19

Ф

- файл options.cfg 29
- файлы данных
 - IBM SPSS Statistics 12
 - импорт и экспорт 12
- файлы данных IBM SPSS Statistics
 - импорт и экспорт 12
- файлы журнала
 - вывод сгенерированного SQL 43
- фиолетовые узлы
 - оптимизация SQL 40
- Функция @SQLFN 51



Напечатано в Дании