

*IBM SPSS Modeler Server 16
Verwaltungs- und
Leistungshandbuch*

IBM

Hinweis

Vor Verwendung dieser Informationen und des darin beschriebenen Produkts sollten die Informationen unter „Bemerkungen“ auf Seite 71 gelesen werden.

Produktinformation

Diese Ausgabe bezieht sich auf Version 16, Release 0, Modifikation 0 von IBM(r) SPSS(r) Modeler und alle nachfolgenden Releases und Modifikationen, bis dieser Hinweis in einer Neuauflage geändert wird.

Diese Veröffentlichung ist eine Übersetzung des Handbuchs *IBM SPSS Modeler Server 16, Administration and Performance Guide*, herausgegeben von International Business Machines Corporation, USA

© Copyright International Business Machines Corporation 2013

Informationen, die nur für bestimmte Länder Gültigkeit haben und für Deutschland, Österreich und die Schweiz nicht zutreffen, wurden in dieser Veröffentlichung im Originaltext übernommen.

Möglicherweise sind nicht alle in dieser Übersetzung aufgeführten Produkte in Deutschland angekündigt und verfügbar; vor Entscheidungen empfiehlt sich der Kontakt mit der zuständigen IBM Geschäftsstelle.

Änderung des Textes bleibt vorbehalten.

Herausgegeben von:
TSC Germany
Kst. 2877
Oktober 2013

Inhaltsverzeichnis

Vorwort	v
--------------------------	----------

Kapitel 1. Informationen zu IBM SPSS

Modeler	1
IBM SPSS Modeler-Produkte	1
IBM SPSS Modeler	1
IBM SPSS Modeler Server	1
IBM SPSS Modeler Administration Console	2
IBM SPSS Modeler Batch	2
IBM SPSS Modeler Solution Publisher	2
IBM SPSS Modeler Server-Adapter für IBM SPSS Collaboration and Deployment Services	2
IBM SPSS Modeler-Editionen	2
IBM SPSS Modeler-Dokumentation	3
SPSS Modeler Professional-Dokumentation	3
SPSS Modeler Premium-Dokumentation	4
Anwendungsbeispiele	5
Ordner "Demos"	5

Kapitel 2. Architektur- und Hardwareempfehlungen **7**

IBM SPSS Modeler - Architektur	7
Architekturbeschreibung	7
Hardwareempfehlungen	9
Temporäre Festplattenspeicher- und Arbeitsspeicheranforderungen	10
Datenzugriff	11
Referenzieren von Datendateien	12
Importieren von IBM SPSS Statistics-Datendateien	12
Installationsanweisungen	13

Kapitel 3. IBM SPSS Modeler - Support **15**

Verbinden mit IBM SPSS Modeler Server	15
Konfigurieren von Single Sign-on	16
Hinzufügen und Bearbeiten der IBM SPSS Modeler Server-Verbindung	17
Suchen nach Servern in IBM SPSS Collaboration and Deployment Services	18
Daten- und Dateisysteme	19
Benutzerauthentifizierung	19
Dateierstellung	19
Unterschiede bei den Ergebnissen	20

Kapitel 4. IBM SPSS Modeler Server-Verwaltung **21**

Starten und Stoppen von IBM SPSS Modeler Server	21
Starten, Stoppen und Statusprüfung unter Windows	21
Starten, Stoppen und Statusprüfung unter UNIX	21
Umgang mit nicht reagierenden Serverprozessen (UNIX-Systeme)	22
Verwaltung	22

Verwenden von IBM SPSS Modeler Administration Console	23
Starten von Modeler Administration Console	23
Konfigurieren des Zugriffs mit Modeler Administration Console	23
SPSS Modeler Server-Verbindungen	24
SPSS Modeler Server-Konfiguration	24
SPSS Modeler Server-Überwachung	29
Verwenden der Datei options.cfg	29
Schließen nicht verwendeter Datenbankverbindungen	30
Verwenden von SSL zur sicheren Datenübertragung	31
Funktionsweise von SSL	31
Schützen der Client/Server- und Server/Server-Kommunikation durch SSL	31

Kapitel 5. Leistungsüberblick **35**

Leistungs- und Optimierungseinstellungen für Ihren Server	35
Leistungs- und Optimierungseinstellungen für Ihren Client	35
Datenbankverwendung und -optimierung	37
SQL-Optimierung	37

Kapitel 6. SQL-Optimierung **39**

Funktionsweise der SQL-Generierung	40
Beispiel für SQL-Generierung	41
Konfigurieren der SQL-Optimierung	42
Vorschau für generiertes SQL	43
Anzeigen von SQL für Modellnuggets	43
Tipps für die Maximierung der SQL-Generierung	43
Knoten zur Unterstützung der SQL-Generierung	44
CLEM-Ausdrücke und Operatoren, die die SQL-Generierung unterstützen	48
Verwenden von SQL-Funktionen in CLEM-Ausdrücken	51
Schreiben von SQL-Abfragen	51

Anhang A. Konfigurieren von Oracle für UNIX-Plattformen **53**

Konfigurieren von Oracle für die SQL-Optimierung	53
--	----

Anhang B. Konfigurieren von UNIX-Startscripts **55**

Einführung	55
Scripts	55
Automatisches Starten und Stoppen von IBM SPSS Modeler Server	55
Manuelles Starten und Stoppen von IBM SPSS Modeler Server	56
Bearbeiten von Scripts	56
Steuern der Berechtigungen für die Dateierstellung	56
IBM SPSS Modeler Server und das Data Access Pack	56
Fehlerbehebung bei der ODBC-Konfiguration	59

Bibliothekspfade	61
Konfigurieren eines Treibers für den Enterprise-Ansichtsknoten	62

Anhang C. Konfigurieren und Ausführen von SPSS Modeler Server als Nicht-Root-Prozess unter UNIX 63

Einführung	63
Konfigurieren als Nicht-Root-Benutzer ohne private Kennwortdatenbank	63
Konfigurieren als Nicht-Root-Benutzer mit einer privaten Kennwortdatenbank	64
Ausführen von SPSS Modeler Server als Nicht-Root-Benutzer	65
Fehlerbehebung bei fehlgeschlagener Benutzerauthentifizierung	65

Anhang D. Konfigurieren und Ausführen von SPSS Modeler Server mit einer privaten Kennwortdatei unter Windows. 67

Einführung	67
Konfigurieren mit einer privaten Kennwortdatenbank.	67

Anhang E. Lastenausgleich mit Server-Clustern 69

Bemerkungen. 71

Marken.	72
-----------------	----

Index 73

Vorwort

IBM® SPSS Modeler ist die auf Unternehmensebene einsetzbare Data-Mining-Workbench von IBM. Mit SPSS Modeler können Unternehmen und Organisationen die Beziehungen zu ihren Kunden bzw. zu den Bürgern durch ein tief greifendes Verständnis der Daten verbessern. Organisationen verwenden die mithilfe von SPSS Modeler gewonnenen Erkenntnisse zur Bindung profitabler Kunden, zur Ermittlung von Cross-Selling-Möglichkeiten, zur Gewinnung neuer Kunden, zur Ermittlung von Betrugsfällen, zur Reduzierung von Risiken und zur Verbesserung der Verfügbarkeit öffentlicher Dienstleistungen.

Die grafische Schnittstelle von SPSS Modeler erleichtert die Anwendung des jeweiligen Fachwissens der Benutzer, was zu leistungsstärkeren Vorhersagemodellen und einer schnelleren Lösungserstellung führt. SPSS Modeler bietet zahlreiche Modellierungsverfahren, beispielsweise Algorithmen für Vorhersage, Klassifizierung, Segmentierung und Assoziationserkennung. Nach der Modellerstellung ermöglicht IBM SPSS Modeler Solution Publisher die unternehmensweite Bereitstellung des Modells für Entscheidungsträger oder in einer Datenbank.

Informationen zu IBM Business Analytics

Die Software IBM Business Analytics liefert umfassende, einheitliche und korrekte Informationen, mit denen Entscheidungsträger die Unternehmensleistung verbessern können. Ein umfassendes Portfolio aus Anwendungen für Business Intelligence, Vorhersageanalyse, Finanz- und Strategiemangement sowie Analysen bietet Ihnen sofort klare und umsetzbare Einblicke in die aktuelle Leistung und gibt Ihnen die Möglichkeit, zukünftige Ergebnisse vorherzusagen. Durch umfassende Branchenlösungen, bewährte Vorgehensweisen und professionellen Service können Unternehmen jeder Größe die Produktivität maximieren, Entscheidungen automatisieren und bessere Ergebnisse erzielen.

Als Teil dieses Portfolios unterstützt IBM SPSS Predictive Analytics-Software Unternehmen dabei, zukünftige Ereignisse vorherzusagen und proaktiv Maßnahmen zu ergreifen, um bessere Geschäftsergebnisse zu erzielen. Kunden aus Wirtschaft, öffentlichem Dienst und staatlichen Lehr- und Forschungseinrichtungen weltweit nutzen IBM SPSS-Technologie als Wettbewerbsvorteil für die Kundengewinnung, Kundenbindung und Erhöhung der Kundenumsätze bei gleichzeitiger Eindämmung der Betrugsmöglichkeiten und Minderung von Risiken. Durch Einbindung von IBM SPSS-Software in ihre täglichen Abläufe wandeln sich Unternehmen zu "Predictive Enterprises", die ihre Entscheidungen auf Geschäftsziele ausrichten und automatisieren können und messbare Wettbewerbsvorteile erzielen. Wenn Sie weitere Informationen wünschen oder Kontakt zu einem Mitarbeiter aufnehmen möchten, besuchen Sie die Seite <http://www.ibm.com/spss>.

Technical Support

Kunden mit Wartungsvertrag können den Technical Support in Anspruch nehmen. Kunden können sich an den Technical Support wenden, wenn sie Hilfe bei der Arbeit mit IBM Produkten oder bei der Installation in einer der unterstützten Hardwareumgebungen benötigen. Den Technical Support erreichen Sie über die Website von IBM unter <http://www.ibm.com/support>. Sie müssen bei der Kontaktaufnahme Ihren Namen, Ihre Organisation und Ihre Supportvereinbarung angeben.

Kapitel 1. Informationen zu IBM SPSS Modeler

IBM SPSS Modeler ist ein Set von Data-Mining-Tools, mit dem Sie auf der Grundlage Ihres Fachwissens schnell und einfach Vorhersagemodelle erstellen und zur Erleichterung der Entscheidungsfindung in die Betriebsabläufe einbinden können. Das Produkt IBM SPSS Modeler, das auf der Grundlage des den Industrienormen entsprechenden Modells CRISP-DM entwickelt wurde, unterstützt den gesamten Data Mining-Prozess, von den Daten bis hin zu besseren Geschäftsergebnissen.

IBM SPSS Modeler bietet eine Vielzahl von Modellbildungsmethoden, die aus dem maschinellen Lernen, der künstlichen Intelligenz und der Statistik stammen. Mit den in der Modellierungspalette verfügbaren Methoden können Sie aus Ihren Daten neue Informationen ableiten und Vorhersagemodelle erstellen. Jede Methode hat ihre speziellen Stärken und eignet sich besonders für bestimmte Problemtypen.

SPSS Modeler kann als Standalone-Produkt oder als Client in Verbindung mit SPSS Modeler Server erworben werden. Außerdem ist eine Reihe von Zusatzoptionen verfügbar, die in den folgenden Abschnitten kurz zusammengefasst werden. Weitere Informationen finden Sie unter <http://www.ibm.com/software/analytics/spss/products/modeler/>.

IBM SPSS Modeler-Produkte

Zur IBM SPSS Modeler-Produktfamilie und der zugehörigen Software gehören folgende Elemente.

- IBM SPSS Modeler
- IBM SPSS Modeler Server
- IBM SPSS Modeler Administration Console
- IBM SPSS Modeler Batch
- IBM SPSS Modeler Solution Publisher
- IBM SPSS Modeler Server-Adapter für IBM SPSS Collaboration and Deployment Services

IBM SPSS Modeler

SPSS Modeler ist eine funktionell in sich abgeschlossene Produktversion, die Sie auf Ihrem PC installieren und ausführen können. Sie können SPSS Modeler im lokalen Modus als Standalone-Produkt oder im verteilten Modus zusammen mit IBM SPSS Modeler Server verwenden, um bei Datasets die Leistung zu verbessern.

Mit SPSS Modeler können Sie schnell und intuitiv genaue Vorhersagemodelle erstellen, und das ohne Programmierung. Mithilfe der speziellen visuellen Benutzerschnittstelle können Sie den Data Mining-Prozess auf einfache Weise visualisieren. Mit der Unterstützung der in das Produkt eingebetteten erweiterten Analyseprozesse können Sie zuvor verborgene Muster und Trends in Ihren Daten aufdecken. Sie können Ergebnisse modellieren und Einblick in die Faktoren gewinnen, die Einfluss auf diese Ergebnisse haben, wodurch Sie in die Lage versetzt werden, Geschäftschancen zu nutzen und Risiken zu mindern.

SPSS Modeler ist in zwei Editionen erhältlich: SPSS Modeler Professional und SPSS Modeler Premium. Weitere Informationen finden Sie unter im Thema „IBM SPSS Modeler-Editionen“ auf Seite 2.

IBM SPSS Modeler Server

SPSS Modeler verwendet eine Client/Server-Architektur zur Verteilung von Anforderungen für ressourcenintensive Vorgänge an leistungsstarke Serversoftware, wodurch bei größeren Datasets eine höhere Leistung erzielt werden kann.

SPSS Modeler Server ist ein separat lizenziertes Produkt, das durchgehend im verteilten Analysemodus auf einem Server-Host in Verbindung mit einer oder mehreren IBM SPSS Modeler-Installationen ausgeführt wird. Auf diese Weise bietet SPSS Modeler Server eine herausragende Leistung bei großen Datensets, da speicherintensive Vorgänge auf dem Server ausgeführt werden können, ohne Daten auf den Client-Computer herunterladen zu müssen. IBM SPSS Modeler Server bietet außerdem Unterstützung für SQL-Optimierung sowie Möglichkeiten zur Modellierung innerhalb der Datenbank, was weitere Vorteile hinsichtlich Leistung und Automatisierung mit sich bringt.

IBM SPSS Modeler Administration Console

Modeler Administration Console ist eine grafische Anwendung zur Verwaltung einer Vielzahl der SPSS Modeler Server-Konfigurationsoptionen, die auch mithilfe einer Optionsdatei konfiguriert werden können. Die Anwendung bietet eine Konsolenbenutzerschnittstelle zur Überwachung und Konfiguration der SPSS Modeler Server-Installationen und steht aktuellen SPSS Modeler Server-Kunden kostenlos zur Verfügung. Die Anwendung kann nur unter Windows installiert werden. Der von ihr verwaltete Server kann jedoch auf einer beliebigen unterstützten Plattform installiert sein.

IBM SPSS Modeler Batch

Das Data-Mining ist zwar in der Regel ein interaktiver Vorgang, es ist jedoch auch möglich, SPSS Modeler über eine Befehlszeile auszuführen, ohne dass die grafische Benutzerschnittstelle verwendet werden muss. Beispielsweise kann es sinnvoll sein, langwierige oder sich wiederholende Aufgaben ohne Eingreifen des Benutzers durchzuführen. SPSS Modeler Batch ist eine spezielle Version des Produkts, die die vollständigen Analysefunktionen von SPSS Modeler ohne Zugriff auf die reguläre Benutzerschnittstelle bietet. Zur Verwendung von SPSS Modeler Batch ist eine SPSS Modeler Server-Lizenz erforderlich.

IBM SPSS Modeler Solution Publisher

SPSS Modeler Solution Publisher ist ein Tool, mit dem Sie eine gepackte Version eines SPSS Modeler-Streams erstellen können, der durch eine externe Runtime-Engine ausgeführt oder in eine externe Anwendung eingebettet werden kann. Auf diese Weise können Sie vollständige SPSS Modeler-Streams für die Verwendung in Umgebungen veröffentlichen und bereitstellen, in denen SPSS Modeler nicht installiert ist. SPSS Modeler Solution Publisher wird als Teil des Diensts für IBM SPSS Collaboration and Deployment Services - Scoring verteilt, für den eine separate Lizenz erforderlich ist. Mit dieser Lizenz erhalten Sie SPSS Modeler Solution Publisher Runtime, womit Sie die veröffentlichten Streams ausführen können.

IBM SPSS Modeler Server-Adapter für IBM SPSS Collaboration and Deployment Services

Für IBM SPSS Collaboration and Deployment Services ist eine Reihe von Adaptern verfügbar, mit denen SPSS Modeler und SPSS Modeler Server mit einem IBM SPSS Collaboration and Deployment Services-Repository interagieren können. Auf diese Weise kann ein im Repository bereitgestellter SPSS Modeler-Stream von mehreren Benutzern gemeinsam verwendet werden. Auch der Zugriff über die Thin-Client-Anwendung IBM SPSS Modeler Advantage ist möglich. Sie installieren den Adapter auf dem System, das als Host für das Repository fungiert.

IBM SPSS Modeler-Editionen

SPSS Modeler ist in den folgenden Editionen erhältlich.

SPSS Modeler Professional

SPSS Modeler Professional bietet sämtliche Tools, die Sie für die Arbeit mit den meisten Typen von strukturierten Daten benötigen, beispielsweise in CRM-Systemen erfasste Verhaltensweisen und Interaktionen, demografische Daten, Kaufverhalten und Umsatzdaten.

SPSS Modeler Premium

SPSS Modeler Premium ist ein separat lizenziertes Produkt, das SPSS Modeler Professional für die Arbeit mit spezialisierten Daten, wie beispielsweise Daten, die für Entitätsanalysen oder soziale Netze verwendet werden, sowie für die Arbeit mit unstrukturierten Textdaten erweitert. SPSS Modeler Premium umfasst die folgenden Komponenten.

IBM SPSS Modeler Entity Analytics fügt den IBM SPSS Modeler-Vorhersageanalysen eine weitere Dimension hinzu. Während bei Vorhersageanalysen versucht wird, zukünftiges Verhalten aus früheren Daten vorherzusagen, liegt der Schwerpunkt bei der Entitätsanalyse auf der Verbesserung von Kohärenz und Konsistenz der aktuellen Daten, indem Identitätskonflikte innerhalb der Datensätze selbst aufgelöst werden. Bei der Identität kann es sich um die Identität einer Person, einer Organisation, eines Objekts oder einer anderen Entität handeln, bei der Unklarheiten bestehen könnten. Die Identitätsauflösung kann in einer Reihe von Bereichen entscheidend sein, darunter Customer Relationship Management, Betrugserkennung, Bekämpfung der Geldwäsche sowie nationale und internationale Sicherheit.

IBM SPSS Modeler Social Network Analysis transformiert Informationen zu Beziehungen in Felder, die das Sozialverhalten von Einzelpersonen und Gruppen charakterisieren. Durch die Verwendung von Daten, die die Beziehungen beschreiben, die sozialen Netzen zugrunde liegen, ermittelt IBM SPSS Modeler Social Network Analysis Führungskräfte in sozialen Netzen, die das Verhalten anderer Personen im Netz beeinflussen. Außerdem können Sie feststellen, welche Personen am meisten durch andere Teilnehmer im Netz beeinflusst werden. Durch die Kombination dieser Ergebnisse mit anderen Maßen können Sie aussagekräftige Profile für Einzelpersonen erstellen, die Sie als Grundlage für Ihre Vorhersagemodelle verwenden können. Modelle, die diese sozialen Informationen berücksichtigen, sind leistungsstärker als Modelle, die dies nicht tun.

IBM SPSS Modeler Text Analytics verwendet hoch entwickelte linguistische Technologien und die Verarbeitung natürlicher Sprache (Natural Language Processing, NLP), um eine schnelle Verarbeitung einer großen Vielfalt an unstrukturierten Textdaten zu ermöglichen, um die Schlüsselkonzepte zu extrahieren und zu ordnen und um diese Konzepte in Kategorien zusammenzufassen. Extrahierte Konzepte und Kategorien können mit bestehenden strukturierten Daten, beispielsweise demografischen Informationen, kombiniert und mithilfe der vollständigen Suite der Data-Mining-Tools von IBM SPSS Modeler auf die Modellierung angewendet werden, um bessere und fokussiertere Entscheidungen zu ermöglichen.

IBM SPSS Modeler-Dokumentation

Eine Dokumentation im OnlinehilfefORMAT finden Sie im Hilfemenü von SPSS Modeler. Diese umfasst die Dokumentation für SPSS Modeler, SPSS Modeler Server und SPSS Modeler Solution Publisher sowie das Anwendungshandbuch und weiteres Material zur Unterstützung.

Die vollständige Dokumentation für die einzelnen Produkte (einschließlich Installationsanweisungen) steht im PDF-Format im Ordner *\Documentation* auf der jeweiligen Produkt-DVD zur Verfügung. Installationsdokumente können auch aus dem Internet unter <http://www-01.ibm.com/support/docview.wss?uid=swg27038316> heruntergeladen werden.

Dokumentation in beiden Formaten steht auch im SPSS Modeler Information Center unter <http://publib.boulder.ibm.com/infocenter/spssmodl/v16r0m0/> zur Verfügung.

SPSS Modeler Professional-Dokumentation

Die SPSS Modeler Professional-Dokumentationssuite (ohne Installationsanweisungen) umfasst folgende Dokumente:

- **IBM SPSS Modeler Benutzerhandbuch.** Allgemeine Einführung in die Verwendung von SPSS Modeler, in der u. a. die Erstellung von Datenstreams, der Umgang mit fehlenden Werten, die Erstellung von CLEM-Ausdrücken, die Arbeit mit Projekten und Berichten sowie das Packen von Streams für die Bereitstellung in IBM SPSS Collaboration and Deployment Services, Predictive Applications (Vorhersageanwendungen) oder IBM SPSS Modeler Advantage beschrieben werden.

- **IBM SPSS Modeler Quellen-, Prozess- und Ausgabeknoten.** Beschreibung aller Knoten, die zum Lesen, zum Verarbeiten und zur Ausgabe von Daten in verschiedenen Formaten verwendet werden. Im Grunde sind sie alle Knoten, mit Ausnahme der Modellierungsknoten.
- **IBM SPSS Modeler Modellierungsknoten.** Beschreibungen sämtlicher für die Erstellung von Data-Mining-Modellen verwendeter Knoten. IBM SPSS Modeler bietet eine Vielzahl von Modellbildungsmethoden, die aus dem maschinellen Lernen, der künstlichen Intelligenz und der Statistik stammen.
- **IBM SPSS Modeler Algorithms Guide.** Beschreibung der mathematischen Grundlagen der in IBM SPSS Modeler verwendeten Modellierungsmethoden. Dieses Handbuch steht nur im PDF-Format zur Verfügung.
- **IBM SPSS Modeler Anwendungshandbuch.** Die Beispiele in diesem Handbuch bieten eine kurze, gezielte Einführung in bestimmte Modellierungsmethoden und -verfahren. Eine Online-Version dieses Handbuchs kann auch über das Hilfemenü aufgerufen werden. Weitere Informationen finden Sie unter im Thema „Anwendungsbeispiele“ auf Seite 5.
- **IBM SPSS Modeler Handbuch für Scripterstellung und Automatisierung.** Informationen zur Automatisierung des Systems über Scripterstellung, einschließlich der Eigenschaften, die zur Bearbeitung von Knoten und Streams verwendet werden können.
- **IBM SPSS Modeler Bereitstellungshandbuch.** Informationen zum Ausführen von IBM SPSS Modeler-Streams und -Szenarios als Schritte bei der Verarbeitung von Jobs im IBM SPSS Collaboration and Deployment Services Deployment Manager.
- **IBM SPSS Modeler CLEF-Entwicklerhandbuch.** CLEF bietet die Möglichkeit, Drittanbieterprogramme, wie Datenverarbeitungsroutinen oder Modellierungsalgorithmen, als Knoten in IBM SPSS Modeler zu integrieren.
- **IBM SPSS Modeler Datenbankinternes Mining.** Informationen darüber, wie Sie Ihre Datenbank dazu einsetzen, die Leistung zu verbessern, und wie Sie die Palette der Analysefunktionen über Drittanbieteralgorithmen erweitern.
- **IBM SPSS Modeler Server Verwaltungs- und Leistungshandbuch.** Informationen zur Konfiguration und Verwaltung von IBM SPSS Modeler Server.
- **IBM SPSS Modeler Administration Console Benutzerhandbuch.** Informationen zur Installation und Nutzung der Konsolenbenutzerschnittstelle zur Überwachung und Konfiguration von IBM SPSS Modeler Server. Die Konsole ist als Plug-in für die Deployment Manager-Anwendung implementiert.
- **IBM SPSS Modeler CRISP-DM Handbuch.** Schritt-für-Schritt-Anleitung für das Data Mining mit SPSS Modeler unter Verwendung der CRISP-DM-Methode.
- **IBM SPSS Modeler Batch Benutzerhandbuch.** Vollständiges Handbuch für die Verwendung von IBM SPSS Modeler im Stapelmodus, einschließlich Details zur Ausführung des Stapelmodus und zu Befehlszeilenargumenten. Dieses Handbuch steht nur im PDF-Format zur Verfügung.

SPSS Modeler Premium-Dokumentation

Die SPSS Modeler Premium-Dokumentationssuite (ohne Installationsanweisungen) umfasst folgende Dokumente:

- **IBM SPSS Modeler Entity Analytics Benutzerhandbuch.** Informationen zur Verwendung von Entitätsanalysen mit SPSS Modeler, unter Behandlung der Repository-Installation und -Konfiguration, Entity Analytics-Knoten und Verwaltungsaufgaben.
- **IBM SPSS Modeler Social Network Analysis User Guide.** Ein Handbuch zur Durchführung einer sozialen Netzanalyse mit SPSS Modeler, einschließlich einer Gruppenanalyse und Diffusionsanalyse.
- **SPSS Modeler Text Analytics Benutzerhandbuch.** Informationen zur Verwendung von Textanalysen mit SPSS Modeler, unter Behandlung der Text Mining-Knoten, der interaktiven Workbench sowie von Vorlagen und anderen Ressourcen.
- **IBM SPSS Modeler Text Analytics Administration Console Benutzerhandbuch.** Informationen zur Installation und Nutzung der Konsolenbenutzerschnittstelle zur Überwachung und Konfiguration von IBM SPSS Modeler Server für die Verwendung mit SPSS Modeler Text Analytics . Die Konsole ist als Plug-in für die Deployment Manager-Anwendung implementiert.

Anwendungsbeispiele

Mit den Data-Mining-Tools in SPSS Modeler kann eine große Bandbreite an geschäfts- und unternehmensbezogenen Problemen gelöst werden; die Anwendungsbeispiele dagegen bieten jeweils eine kurze, gezielte Einführung in spezielle Modellierungsmethoden und -verfahren. Die hier verwendeten Datasets sind viel kleiner als die großen Datenbestände, die von einigen Data-Mining-Experten verwaltet werden müssen, die zugrunde liegenden Konzepte und Methoden sollten sich jedoch auch auf reale Anwendungen übertragen lassen.

Sie können auf die Beispiele zugreifen, indem Sie im Menü "Hilfe" in SPSS Modeler auf die Option **Anwendungsbeispiele** klicken. Die Datendateien und Beispielstreams wurden im Ordner *Demos*, einem Unterordner des Produktinstallationsverzeichnis, installiert. Weitere Informationen finden Sie unter im Thema „Ordner "Demos"“.

Beispiele für die Datenbankmodellierung. Die Beispiele finden Sie im IBM SPSS Modeler-Handbuch zum datenbankinternen Mining.

Scriptbeispiele. Die Beispiele finden Sie im *IBM SPSS Modeler Handbuch für Scripterstellung und Automatisierung*.

Ordner "Demos"

Die in den Anwendungsbeispielen verwendeten Datendateien und Beispielstreams wurden im Ordner *Demos*, einem Unterordner des Produktinstallationsverzeichnis, installiert. Auf diesen Ordner können Sie auch über die Programmgruppe IBM SPSS Modeler im Windows-Startmenü oder durch Klicken auf *Demos* in der Liste der zuletzt angezeigten Verzeichnisse im Dialogfeld "Datei öffnen" zugreifen.

Kapitel 2. Architektur- und Hardwareempfehlungen

IBM SPSS Modeler - Architektur

In diesem Kapitel wird die Architektur von IBM SPSS Modeler Server beschrieben, inklusive der Server-Software, der Client-Software und der Datenbank. Es umfasst Informationen zum Design von IBM SPSS Modeler Server, das auf optimale Leistung ausgerichtet ist, sowie Empfehlungen zur erforderlichen Hardware, um diese Leistung voll ausschöpfen zu können. Das Kapitel wird mit einem Abschnitt über den Datenzugriff abgeschlossen. Hier erfahren Sie, wo die erforderlichen ODBC-Treiber eingerichtet werden müssen.

Architekturbeschreibung

IBM SPSS Modeler Server verwendet eine verteilte Architektur mit drei Schichten. Softwareoperationen werden zwischen den Client- und Server-Computern aufgeteilt. Die Vorteile von IBM SPSS Modeler Server gegenüber der Standalone-Version IBM SPSS Modeler, vor allem beim Verarbeiten großer Datasets, sind zahlreich:

- IBM SPSS Modeler Server kann nicht nur unter Windows, sondern auch unter UNIX ausgeführt werden, sodass Sie eine größere Flexibilität hinsichtlich des Installationsortes haben. Sie können auf jeder Plattform einen schnelleren, größeren Server-Computer für Data-Mining-Prozesse verwenden.
- IBM SPSS Modeler Server wurde vorwiegend für Schnelligkeit optimiert. Wenn Vorgänge nicht an die Datenbank übertragen werden können, speichert IBM SPSS Modeler Server die Zwischenergebnisse als temporäre Dateien auf der Festplatte und nicht im Arbeitsspeicher. Da Server normalerweise über ausreichend Festplattenspeicher verfügen, kann IBM SPSS Modeler Server Sortier-, Zusammenführungs- und Aggregationsvorgänge mit sehr großen Datasets durchführen.
- Mit der Client/Server-Architektur können Sie die Data-Mining-Prozesse in Ihrem Unternehmen zentralisieren. Auf diese Weise können Sie die Rolle des Data-Mining in Ihren Geschäftsprozessen formalisieren.
- Mithilfe von Verwaltungstools wie IBM SPSS Modeler Administration Console (in IBM SPSS Modeler Server enthalten) sowie IBM SPSS Collaboration and Deployment Services (separat erhältlich) können Sie Data-Mining-Prozesse überwachen und sicherstellen, dass ausreichend Computerressourcen verfügbar sind. Mit IBM SPSS Collaboration and Deployment Services können Sie bestimmte Data-Mining-Aufgaben automatisieren, den Zugriff auf Datenmodelle verwalten und Ergebnisse unternehmensweit verfügbar machen.

Die Komponenten der verteilten Architektur von IBM SPSS Modeler sind in der Grafik "IBM SPSS Modeler Server-Architektur" dargestellt.

- **IBM SPSS Modeler.** Die Client-Software wird auf dem Computer des Endbenutzers installiert. Sie stellt die Benutzerschnittstelle zur Verfügung und zeigt Data-Mining-Ergebnisse an. Der Client ist eine vollständige Installation der IBM SPSS Modeler-Software. Wird der Client jedoch für die verteilte Analyse mit IBM SPSS Modeler Server verbunden, wird seine Ausführungseingabe inaktiviert. IBM SPSS Modeler kann nur unter Windows ausgeführt werden.
- **IBM SPSS Modeler Server.** Die Server-Software wird auf einem Server-Computer installiert, der Netzkonnektivität sowohl zu IBM SPSS Modeler-Instanzen als auch zur Datenbank aufweist. IBM SPSS Modeler Server wird als Service (unter Windows) oder als Dämonprozess (unter UNIX) ausgeführt und wartet auf die Verbindung von Clients. Er verarbeitet die Ausführung von Streams und Scripts, die mit IBM SPSS Modeler erstellt wurden.
- **Datenbankserver.** Der Datenbankserver könnte ein Livedaten-Warehouse (z. B. Oracle auf einem großen UNIX-Server) oder, um die Auswirkungen auf andere Betriebssysteme zu verringern, ein Datamart auf einem lokalen Server bzw. einem Abteilungsserver (z. B. SQL Server unter Windows) sein.

IBM SPSS Modeler Server-Architektur

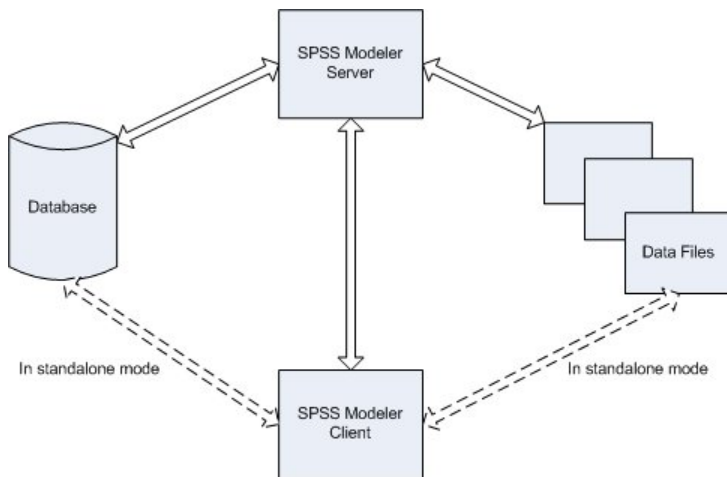


Abbildung 1. IBM SPSS Modeler - Architektur

Bei der verteilten Architektur findet ein Großteil der Verarbeitung auf dem Server-Computer statt. Wenn der Endbenutzer einen Stream ausführt, sendet IBM SPSS Modeler eine Beschreibung des Streams an den Server. Der Server ermittelt, welche Vorgänge in SQL ausgeführt werden können, und erstellt die passenden Anfragen. Diese Anfragen werden in der Datenbank ausgeführt und die Ergebnisse werden an den Server übergeben, der die Verarbeitung der Daten übernimmt, die nicht mithilfe von SQL ausgedrückt werden können. Nachdem die Verarbeitung abgeschlossen ist, werden nur relevante Ergebnisse an den Client zurückgegeben.

Falls erforderlich, kann IBM SPSS Modeler Server alle IBM SPSS Modeler-Vorgänge außerhalb der Datenbank durchführen. Die Verwendung von Arbeitsspeicher und Festplattenspeicher für die Speicherung von Daten für die Verarbeitung wird automatisch im Gleichgewicht gehalten. Durch diesen Prozess ist IBM SPSS Modeler Server vollständig kompatibel mit Flatfiles.

Außerdem steht Lastenausgleich durch die Verwendung von Server-Clustern für die Verarbeitung zur Verfügung. Clustering ist in IBM SPSS Collaboration and Deployment Services ab Version 3.5 über das Plug-in Coordinator of Processes verfügbar. Weitere Informationen finden Sie in Anhang E, „Lastenausgleich mit Server-Clustern“, auf Seite 69. Die Verbindung mit einem im Coordinator of Processes verwalteten Server oder Cluster ist direkt über das IBM SPSS Modeler-Dialogfeld "Anmelden beim Server" möglich. Weitere Informationen finden Sie im Thema „Verbinden mit IBM SPSS Modeler Server“ auf Seite 15.

Standalone-Client

IBM SPSS Modeler kann auch für die Ausführung als eigenständige Desktopanwendung konfiguriert werden, wie in der unten stehenden Grafik gezeigt. Weitere Informationen finden Sie unter Kapitel 3, „IBM SPSS Modeler - Support“, auf Seite 15.

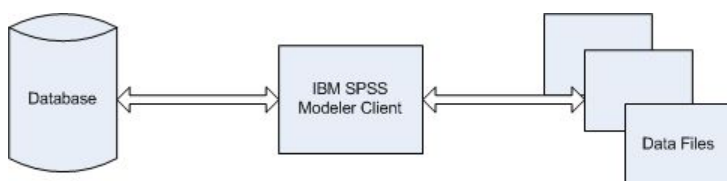


Abbildung 2. IBM SPSS Modeler (Standalone)

Hardwareempfehlungen

Bei der Planung Ihrer IBM SPSS Modeler Server-Installation sollten Sie auch die Hardware berücksichtigen, die Sie verwenden werden. Auch wenn das Design von IBM SPSS Modeler Server auf Geschwindigkeit ausgerichtet ist, können Sie die Effizienz der Anwendung nur mithilfe der richtigen Hardware für Ihre Data-Mining-Aufgaben vollständig ausschöpfen. Ein Aufrüsten der Hardware ist häufig die einfachste und wirtschaftlichste Lösung zum Verbessern der Systemleistung.

Dedizierter Server. Installieren Sie IBM SPSS Modeler Server auf einem dedizierten Server-Computer, dessen Ressourcen allein der Server-Software zur Verfügung stehen. Dies betrifft auch alle Datenbanken, auf die IBM SPSS Modeler Server zugreift. Vor allem Modellerstellungsvorgänge sind besonders ressourcenintensiv und können wesentlich schneller durchgeführt werden, wenn keine Ressourcen für andere Anwendungen abgezweigt werden.

Hinweis: Obwohl die Installation von IBM SPSS Modeler Server und der Datenbank auf demselben Computer die Übertragungszeit zwischen diesen beiden Komponenten aufgrund der verringerten Netzlast reduzieren kann, ist es in den meisten Fällen sinnvoll, Server und Datenbank auf getrennten Computern zu konfigurieren, um Ressourcenkonflikte zu vermeiden. Mit einer schnellen Verbindung zwischen den beiden Komponenten können Sie die Kosten der Datenübertragung minimieren.

Prozessoren. Die Anzahl an Prozessoren im Computer sollte nicht unter der Anzahl an gleichzeitigen Aufgaben (gleichzeitigen Streams) liegen, die regelmäßig ausgeführt werden müssen. Im Allgemeinen gilt: Je mehr Prozessoren, desto besser.

- Eine einzelne Instanz von IBM SPSS Modeler Server akzeptiert Verbindungen von mehreren Clients (Benutzern). Jede Clientverbindung kann mehrere Streamausführungen starten. Auf einem Server können daher stets mehrere Ausführungsaufgaben gleichzeitig laufen.
- Als Faustregel gehen Sie von einem Prozessor für einen bis zwei Benutzer, zwei Prozessoren für bis zu vier Benutzer und acht Prozessoren für bis zu acht Benutzer aus. Je nach Arbeitsanforderung fügen Sie darüber hinaus je einen weiteren Prozessor für alle zwei bis vier Benutzer hinzu.
- Der Anteil der Verarbeitung, der durch die SQL-Optimierung an die Datenbank übergeben werden kann, bestimmt, ob eine CPU von zwei oder mehr Benutzern gemeinsam verwendet werden kann, bei einem nur geringen Leistungsverlust.
- Multithreading-Fähigkeiten ermöglichen, dass eine Aufgabe mehrere Prozessoren verwenden kann, sodass ein Aufstocken der Prozessoren die Leistung auch in Fällen verbessert, in denen jeweils nur eine Aufgabe gleichzeitig durchgeführt wird. Im Allgemeinen wird das Multithreading für die C5.0-Modellerstellung und für bestimmte Datenvorbereitungsvorgänge (Sortieren, Aggregieren und Zusammenführen) verwendet.

64-Bit-Plattformen. Wenn Sie Modelle mit großen Datenmengen erstellen oder verarbeiten möchten, verwenden Sie 64-Bit-Solaris, -Windows oder -Linux als Plattform für IBM SPSS Modeler Server und verwenden Sie die maximal mögliche Menge an RAM für diesen Computer. Bei großen Datensets schöpft der Server die verfügbaren Speichermengen (die den Einschränkungen der 32-Bit-Plattform unterliegen) für Prozesse schnell aus, sodass Daten auf die Festplatte geschrieben werden müssen, was zu beträchtlichen Zeitverlusten bei der Ausführung führt. 64-Bit-Serverimplementierungen können zusätzlichen Arbeitsspeicher nutzen. Es werden jedoch mindestens 8 GB empfohlen. 64-Bit-Unterstützung ist für Solaris-, Windows- und Linux-Plattformen verfügbar.

Langfristige Anforderungen. Sofern möglich, stellen Sie sicher, dass die Serverhardware hinsichtlich Arbeitsspeicher und CPUs erweiterbar ist, um später einer höheren Auslastung (z. B. einer größeren Anzahl an gleichzeitigen Benutzern oder höheren Anforderungen an die Verarbeitung) sowie erweiterten Multithreading-Fähigkeiten von nachfolgenden IBM SPSS Modeler Server-Versionen gerecht werden zu können.

Temporäre Festplattenspeicher- und Arbeitsspeicheranforderungen

IBM SPSS Modeler Server verwendet für die Verarbeitung von großen Datenmengen temporären Festplattenspeicher. Die Menge an erforderlichem Festplattenspeicher hängt von der verarbeiteten Datenart und -menge sowie von den jeweils durchgeführten Vorgängen ab. Die Datenmenge verhält sich proportional zur Anzahl an Zeilen *und* Spalten. Je mehr Zeilen und Spalten verarbeitet werden, desto mehr Festplattenspeicher ist erforderlich.

In diesem Abschnitt werden die Umstände beschrieben, unter denen temporärer Festplattenspeicher und zusätzlicher Arbeitsspeicher erforderlich sind und wie die erforderliche Menge ermittelt wird. Beachten Sie, dass in diesem Abschnitt nicht die Anforderungen an temporären Festplattenspeicher beschrieben werden, die in einer Datenbank auftreten, denn diese Anforderungen hängen von der jeweiligen Datenbank ab.

Bedingungen, die temporären Festplattenspeicher erfordern

Die leistungsstarke SQL-Optimierungsfunktion von IBM SPSS Modeler Server ermöglicht die Verarbeitung innerhalb der Datenbank (anstelle des Servers). Unter den folgenden Umständen kann die SQL-Optimierung jedoch nicht verwendet werden:

- Die zu verarbeitenden Daten befinden sich in einer Flatfile und nicht in einer Datenbank.
- Die SQL-Optimierung ist inaktiviert.
- Die Verarbeitung kann mit SQL nicht optimiert werden.

Wenn die SQL-Optimierung nicht verwendet werden kann, erstellen die folgenden Datenbearbeitungsknoten und CLEM-Funktionen auf der Festplatte temporäre Kopien von einigen oder allen Daten. Wenn die auf Ihrer Site verwendeten Streams diese Verarbeitungsbefehle oder -Funktionen enthalten, müssen Sie möglicherweise zusätzlichen Festplattenspeicher auf Ihrem Server bereitstellen.

- Aggregatknotten
- Duplikatknotten
- Klassierknotten
- Mergeknotten bei Verwendung der Option "Merge nach Schlüssel"
- Jeder Modellierungsknoten
- Sortierknotten
- Tabellenausgabeknotten
- @OFFSET-Funktionen, bei denen die Suchbedingung die Funktion @THIS verwendet.
- Jede @-Funktion wie @MIN, @MAX und @AVE, in denen der Offset-Parameter berechnet wird.

Berechnen der Menge an erforderlichem Festplattenspeicher

Im Allgemeinen muss IBM SPSS Modeler Server in der Lage sein, eine temporäre Datei zu erstellen, die mindestens *dreimal so groß* ist, wie das ursprüngliche Dataset. Wenn die Datendatei beispielsweise 2 GB groß ist und die SQL-Generierung nicht verwendet wird, benötigt IBM SPSS Modeler Server 6 GB Festplattenspeicher, um die Daten zu verarbeiten. Da jedes gleichzeitig angemeldete Benutzerkonto eigene temporäre Dateien erstellt, müssen Sie den Festplattenspeicher für jeden dieser Benutzer entsprechend anpassen.

Wenn Sie feststellen, dass Ihre Site häufig große temporäre Dateien verwendet, können Sie für die temporären Dateien von IBM SPSS Modeler ein separates Dateisystem auf einer separaten Festplatte einrichten. Die besten Ergebnisse, um die Festplattenvorgänge zu beschleunigen, erzielen Sie mit einem RAID 0 oder Striped Data Set, das mehrere physische Festplatten umfasst, idealerweise mit jeder Festplatte im Striped-Dateisystem auf einem separaten Festplattencontroller.

Arbeitsspeicheranforderungen

Für die meisten Vorgänge, die nicht in der Datenbank verarbeitet werden können, speichert IBM SPSS Modeler Server die Zwischenergebnisse als temporäre Dateien auf der Festplatte und nicht im Arbeitsspeicher (RAM). Im Falle von Modellierungsknoten wird jedoch der Arbeitsspeicher verwendet, wenn

möglich. Die Knoten "Neuronales Netz", "Kohonen" und "K-Means" benötigen eine große Menge an Arbeitsspeicher. Wenn diese Knoten häufig auf Ihrer Site verwendet werden, rüsten Sie gegebenenfalls den Arbeitsspeicher Ihres Servers auf.

Im Allgemeinen kann die Menge an erforderlichem Arbeitsspeicher wie folgt geschätzt werden:
(Anzahl_an_Datensätzen * Anzahl_an_Zellen_pro_Datensatz) * Anzahl_an_Byte_pro_Zelle.

Hierbei kann Anzahl_an_Zellen_pro_Datensatz sehr groß werden, wenn nominale Felder verwendet werden.

Die aktuell gültigen RAM-Anforderungen finden Sie im Serverinstallationshandbuch im Abschnitt zu den Systemanforderungen. Ab vier gleichzeitigen Benutzern wird noch mehr RAM empfohlen. Der Arbeitsspeicher muss unter gleichzeitig durchgeführten Aufgaben aufgeteilt werden, daher sollten Sie entsprechend aufrüsten. Grundsätzlich ist das Aufrüsten von Arbeitsspeicher einer der kostengünstigsten Wege, die Systemleistung zu verbessern.

Datenzugriff

Um in einer Datenbank zu lesen oder in ihr zu schreiben, muss eine ODBC-Datenquelle für die entsprechende Datenbank mit den erforderlichen Lese- und Schreibberechtigungen installiert und konfiguriert sein. Das IBM SPSS Data Access Pack umfasst eine Reihe von ODBC-Treibern, die zu diesem Zweck verwendet werden können. Diese Treiber stehen auf der IBM SPSS Data Access Pack-DVD oder auf der Download-Site zur Verfügung. Wenn Sie Fragen zur Erstellung oder Einstellung von Berechtigungen für ODBC-Datenquellen haben, wenden Sie sich an Ihren Datenbankadministrator.

Unterstützte ODBC-Treiber

Neueste Informationen zu Datenbanken und ODBC-Treibern, die für die Verwendung mit IBM SPSS Modeler 16 getestet wurden und unterstützt werden, finden Sie in den Produktkompatibilitätsdiagrammen auf der unternehmensweiten Support-Site unter <http://www.ibm.com/support>.

Installationsort der Treiber

Beachten Sie, dass die ODBC-Treiber auf jedem Computer installiert und konfiguriert werden müssen, auf dem eine Verarbeitung erfolgt.

- Wenn Sie IBM SPSS Modeler im lokalen Modus (Standalone-Modus) ausführen, müssen die Treiber auf dem lokalen Computer installiert sein.
- Wenn Sie IBM SPSS Modeler im verteilten Modus mit einer fernen IBM SPSS Modeler Server-Instanz ausführen, müssen die ODBC-Treiber auf dem Computer installiert sein, auf dem IBM SPSS Modeler Server installiert ist. Beachten Sie bei IBM SPSS Modeler Server auf UNIX-Systemen auch "Konfiguration von ODBC-Treibern auf UNIX-Systemen" weiter hinten in diesem Abschnitt.
- Wenn Sie von IBM SPSS Modeler und IBM SPSS Modeler Server auf dieselben Datenquellen zugreifen müssen, müssen die ODBC-Treiber auf beiden Computern installiert sein.
- Wenn Sie IBM SPSS Modeler über Terminaldienste ausführen, müssen die ODBC-Treiber auf dem Terminaldiensteserver installiert sein, auf dem Sie IBM SPSS Modeler installiert haben.

Wichtig: Wenn Sie IBM SPSS Modeler Server unter UNIX zum Zugriff auf eine Teradata-Datenbank verwenden und den von Teradata bereitgestellten ODBC-Treiber verwenden, müssen Sie den ODBC-Treibermanager verwenden, der mit diesem Treiber installiert wurde. (**Hinweis:** Sie müssen diese Änderungen nicht vornehmen, wenn Sie den SDAP-Teradata-Treiber verwenden.) Um diese Änderung an IBM SPSS Modeler Server vorzunehmen, geben Sie für ODBC_DRIVER_MANAGER_PATH einen Wert in der Nähe des oberen Bereichs des Scripts modelersrv.sh ein, wo dies durch die Kommentare angegeben wurde. Diese Umgebungsvariable muss auf den Speicherort des ODBC-Treibermanagers eingestellt werden, der mit dem Teradata ODBC-Treiber ausgeliefert wird (/usr/odbc/lib in einer Standardinstallation eines Teradata-ODBC-Treibers). Sie müssen IBM SPSS Modeler Server neu starten, damit die Änderung wirksam wird.

Weitere Informationen zu den IBM SPSS Modeler Server-Plattformen, die Teradata-Zugriff unterstützen, sowie über die unterstützte Teradata ODBC-Treiberversion finden Sie auf der unternehmensweiten Support-Site unter <http://www.ibm.com/support>.

Konfiguration von ODBC-Treibern auf UNIX-Systemen

Standardmäßig ist der DataDirect-Treibermanager nicht für IBM SPSS Modeler Server auf UNIX-Systemen konfiguriert. Geben Sie folgende Befehle ein, um UNIX so zu konfigurieren, dass der DataDirect-Treibermanager geladen wird:

```
cd Modeler Server-Installationsverzeichnis/bin
rm -f libspssodbc.so
ln -s libspssodbc_datadirect.so libspssodbc.so
```

Dadurch wird die Standardverknüpfung entfernt und eine Verknüpfung zum DataDirect-Treibermanager erstellt.

Hinweis: Die oben genannten Regeln betreffen ausdrücklich den Zugriff auf Daten in einer Datenbank. Andere Dateizugriffe, wie das Öffnen und Speichern von Stream-, Projekt-, Modell-, Knoten-, PMML-, Ausgabe- und Scriptdateien, finden stets auf dem Client statt und werden stets unter Berücksichtigung des Dateisystems auf dem Client-Computer angegeben. Ferner legt der Befehl "Verzeichnis festlegen" in IBM SPSS Modeler das Arbeitsverzeichnis von *lokalen* Clientobjekten (z. B. Streams) fest, er hat jedoch keinerlei Auswirkung auf das Arbeitsverzeichnis des Servers.

UNIX. Informationen zum Konfigurieren von IBM SPSS Modeler Server unter UNIX, um mit der IBM SPSS Statistics-Datenzugriffstechnologie zu arbeiten, finden Sie unter Anhang B, „Konfigurieren von UNIX-Startscripts“, auf Seite 55.

Referenzieren von Datendateien

Windows. Wenn Sie Daten auf demselben Computer wie IBM SPSS Modeler Server speichern, wird empfohlen, den Pfad zu den Daten aus der Perspektive des Server-Computers anzugeben (z. B. *C:\ServerData\Sales 1998.csv*). Die Leistung ist besser, wenn die Datei nicht über das Netz gesucht werden muss.

Sind die Daten auf einem anderen Host gespeichert, wird die Verwendung von UNC-Dateiverweisen empfohlen (z. B. *\\Mein Datenserver\ServerData\Sales 1998.csv*). Beachten Sie, dass UNC-Namen nur funktionieren, wenn der Pfad den Namen der gemeinsamen Netzressource enthält. Der referenzierende Computer muss die Berechtigung zum Lesen der angegebenen Datei haben. Wenn Sie häufig vom verteilten in den lokalen Analysemodus wechseln, verwenden Sie UNC-Dateiverweise, da diese unabhängig vom Modus immer funktionieren.

UNIX. Zum Referenzieren von Daten auf einem UNIX-Server verwenden Sie die vollständige Dateispezifizierung und normale Schrägstriche (z. B. */public/data/ServerData/Umsatz1998.csv*). Vermeiden Sie im UNIX-Verzeichnis und in Dateinamen für Daten, die mit IBM SPSS Modeler Server verwendet werden, den umgekehrten Schrägstrich. Es spielt keine Rolle, ob eine Textdatei im UNIX- oder DOS-Format vorhanden ist, da beide automatisch verarbeitet werden.

Importieren von IBM SPSS Statistics-Datendateien

Wenn Sie auf Ihrer Site IBM SPSS Statistics Server ausführen, kann es vorkommen, dass Benutzer IBM SPSS Statistics-Daten im verteilten Modus importieren oder exportieren möchten. Beachten Sie, dass IBM SPSS Modeler im verteilten Modus das Dateisystem des Servers verwendet. Der IBM SPSS Statistics Client arbeitet auf dieselbe Weise. Zum Importieren und Exportieren zwischen den beiden Anwendungen müssen die beiden Clients in demselben Modus arbeiten. Ist dies nicht der Fall, weichen die Ansichten der Dateisysteme voneinander ab und die gemeinsame Nutzung der Dateien ist nicht möglich. Die IBM

SPSS Statistics-Knoten von IBM SPSS Modeler können den IBM SPSS Statistics Client automatisch starten, die Benutzer müssen jedoch zunächst sicherstellen, dass der IBM SPSS Statistics Client und IBM SPSS Modeler in demselben Modus arbeiten.

Installationsanweisungen

Informationen zur Installation von IBM SPSS Modeler Server finden Sie unter `\documentation\installation\` auf der Server-CD. Weitere Dokumente sind für Windows und UNIX verfügbar.

Vollständige Informationen zum Installieren und Arbeiten mit IBM SPSS Modeler finden Sie auf der Client-CD.

Kapitel 3. IBM SPSS Modeler - Support

Dieser Abschnitt wurde für Administratoren und Help-Desk-Mitarbeiter geschrieben, die Support für die Benutzer von IBM SPSS Modeler anbieten. Die folgenden Themen werden beschrieben:

- Anmeldung bei IBM SPSS Modeler Server (oder Standalone-Ausführung durch Trennen der Verbindung zu einem Server)
- Vom Benutzer benötigte Daten- und Dateisysteme
- Benutzerkonten und Dateiberechtigungen für IBM SPSS Modeler Server
- Unterschiede bei den Ergebnissen, die Benutzer beim Wechsel zwischen IBM SPSS Modeler Server und IBM SPSS Modeler sehen

Verbinden mit IBM SPSS Modeler Server

IBM SPSS Modeler kann als eigenständige Anwendung oder als Client ausgeführt werden, der direkt mit IBM SPSS Modeler Server oder über das Plug-in Coordinator of Processes von IBM SPSS Collaboration and Deployment Services mit einer IBM SPSS Modeler Server-Instanz oder einem Server-Cluster verbunden ist. Der aktuelle Verbindungsstatus wird unten links im IBM SPSS Modeler-Fenster angezeigt.

Wenn Sie eine Verbindung zu einem Server herstellen möchten, können Sie den Namen des Servers, mit dem eine Verbindung hergestellt werden soll, manuell eingeben oder einen zuvor definierten Namen auswählen. Wenn Sie IBM SPSS Collaboration and Deployment Services verwenden, können Sie im Dialogfeld für die Anmeldung beim Server eine Liste von Servern bzw. Server-Clustern durchsuchen. Die Möglichkeit, die auf einem Netz ausgeführten Statistics-Dienste zu durchsuchen, wird über den Coordinator of Processes bereitgestellt.

So stellen Sie eine Verbindung mit einem Server her:

1. Klicken Sie im Menü "Extras" auf die Option **Anmelden beim Server**. Das Dialogfeld "Anmelden beim Server" wird geöffnet. Alternativ können Sie auf den Bereich des Verbindungsstatus im IBM SPSS Modeler-Fenster doppelklicken.
2. Legen Sie in diesem Dialogfeld die Optionen zum Verbinden mit dem lokalen Server-Computer fest oder wählen Sie eine Verbindung in der Tabelle aus.
 - Klicken Sie auf **Hinzufügen** bzw. **Bearbeiten**, um eine Verbindung hinzuzufügen bzw. zu bearbeiten. Weitere Informationen finden Sie unter im Thema „Hinzufügen und Bearbeiten der IBM SPSS Modeler Server-Verbindung“ auf Seite 17.
 - Klicken Sie auf **Suche**, um auf einen Server bzw. Server-Cluster in Coordinator of Processes zuzugreifen. Weitere Informationen finden Sie unter im Thema „Suchen nach Servern in IBM SPSS Collaboration and Deployment Services“ auf Seite 18.

Servertabelle. Diese Tabelle enthält die Menge der definierten Serververbindungen. In der Tabelle werden die Standardverbindung, der Servername sowie die Beschreibung und Portnummer angegeben. Sie können manuell eine neue Verbindung hinzufügen sowie eine bestehende Verbindung auswählen bzw. danach suchen. Um einen bestimmten Server als Standardverbindung einzurichten, aktivieren Sie in der Tabelle für die Verbindung das Kontrollkästchen in der Spalte "Standard".

Standarddatenpfad. Geben Sie einen Pfad an, der für Daten auf dem Server-Computer verwendet wird. Mit der Auslassungsschaltfläche (...) wechseln Sie zum gewünschten Verzeichnis.

Berechtigungs-nachweise einstellen. Lassen Sie dieses Kontrollkästchen unausgewählt, um die Funktion für **Single Sign-on** zu aktivieren. Diese versucht, Sie mithilfe Ihres lokalen Benutzernamens und Kennworts beim Server anzumelden. Falls ein Single Sign-on nicht möglich ist oder Sie das Kontrollkästchen zur Inaktivierung des Single Sign-on aktivieren (z. B. zur Anmeldung an einem Administratorkonto), wird ein weiteres Fenster angezeigt, in dem Sie aufgefordert werden, Ihre Berechtigungs-nachweise einzugeben.

Benutzer-ID. Geben Sie den Benutzernamen ein, mit dem die Anmeldung beim Server erfolgen soll.

Kennwort. Geben Sie das Kennwort ein, das zum angegebenen Benutzernamen gehört.

Domäne. Geben Sie die Domäne an, mit der die Anmeldung beim Server erfolgen soll. Ein Domänenname ist nur dann erforderlich, wenn sich der Server-Computer in einer anderen Windows-Domäne befindet als der Client-Computer.

3. Klicken Sie auf **OK**, um die Verbindung herzustellen.

So trennen Sie eine Verbindung mit einem Server:

1. Klicken Sie im Menü "Extras" auf die Option **Anmelden beim Server**. Das Dialogfeld "Anmelden beim Server" wird geöffnet. Alternativ können Sie auf den Bereich des Verbindungsstatus im IBM SPSS Modeler-Fenster doppelklicken.
2. Wählen Sie im Dialogfeld den lokalen Server aus und klicken Sie auf **OK**.

Konfigurieren von Single Sign-on

Sie können mit Single Sign-on eine Verbindung zu einer Instanz von IBM SPSS Modeler Server herstellen, die auf einer beliebigen unterstützten Plattform ausgeführt wird. Wenn Sie eine Verbindung mit Single Sign-on herstellen wollen, müssen Sie zunächst Ihren IBM SPSS Modeler-Server und die Client-Computer konfigurieren.

Wenn Sie Single Sign-On verwenden, um die Verbindung sowohl zu IBM SPSS Modeler Server als auch zu IBM SPSS Collaboration and Deployment Services herzustellen, müssen Sie zuerst die Verbindung zu IBM SPSS Collaboration and Deployment Services herstellen, bevor Sie die Verbindung zu IBM SPSS Modeler herstellen.

Anmerkung: Bevor Sie Ihren IBM SPSS Modeler-Server und die Client-Computer für Single Sign-on konfigurieren, müssen Sie sicherstellen, dass die Computer Zugriff auf den Domänencontroller-Server haben.

Konfigurieren von Single Sign-on unter Windows

So konfigurieren Sie IBM SPSS Modeler Server:

1. Stellen Sie sicher, dass die Windows-Servermaschine zur Active Directory-Domäne gehört.
2. Suchen Sie im IBM SPSS Modeler Server-Installationsverzeichnis den Ordner config.
3. Erstellen Sie im Ordner config den Unterordner sso.
4. Erstellen Sie im Ordner sso die Datei krb5.conf. Anweisungen zum Erstellen der Datei krb5.conf finden Sie unter http://web.mit.edu/kerberos/krb5-current/doc/admin/conf_files/krb5_conf.html. Im Folgenden finden Sie ein Beispiel für eine Datei krb5.conf:

```
[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
```

So konfigurieren Sie den IBM SPSS Modeler-Client:

1. Stellen Sie sicher, dass die lokale Windows-Maschine, auf der IBM SPSS Modeler ausgeführt wird, zur Active Directory-Domäne gehört.
2. Fügen Sie den Domänenbenutzer als Administrator auf der lokalen Maschine hinzu.
3. Aktivieren Sie Windows für den Zugriff auf den TGT-Sitzungsschlüssel:

- a. Klicken Sie im Menü **Start** auf **Ausführen**.
 - b. Geben Sie `regedit` ein und klicken Sie auf **OK**, um den **Registrierungseditor** zu öffnen.
 - c. Navigieren Sie zu der Registry-Position für das Betriebssystem der lokalen Maschine:
 - Unter Windows XP: `Computer\HKEY_LOCAL_MACHINE\System\CurrentControlSet\Control\Lsa\Kerberos`
 - Unter Windows Vista oder Windows 7: `Computer\HKEY_LOCAL_MACHINE\System\CurrentControlSet\Control\Lsa\Kerberos\Parameters`
 - d. Klicken Sie mit der rechten Maustaste auf den Ordner und wählen Sie **Neu > DWORD** aus. Der Name des neuen Wertes sollte `allowtgtsessionkey` lauten.
 - e. Setzen Sie den Wert für `allowtgtsessionkey` auf den Hexadezimalwert 1, also `0x0000001`.
 - f. Schließen Sie den **Registrierungseditor**.
 - g. Wenn das Betriebssystem der lokalen Maschine Windows Vista oder Windows 7 ist, müssen Sie die Datei `kinit.exe` ausführen, die sich im Verzeichnis `<IBM SPSS Modeler Installationsverzeichnis>\jre\bin` befindet.
4. Erstellen Sie im Ordner `config` des IBM SPSS Modeler-Installationsverzeichnisses den Ordner `sso`.
 5. Kopieren Sie die Datei `krb5.conf` vom Server in den Ordner `sso`.
 6. Starten Sie die lokale Maschine und die Servermaschine erneut.

Konfigurieren von Single Sign-on unter UNIX

Zum Konfigurieren von Single Sign-On für UNIX-Servermaschinen können Sie die UNIX-Maschine zur Windows Active Directory-Domäne hinzufügen und anschließend den Anweisungen zur Konfiguration von Single Sign-on unter Windows folgen. Weitere Informationen finden Sie im Thema „Konfigurieren von Single Sign-on unter Windows“ auf Seite 16. Alternativ können Sie die folgenden Schritte ausführen:

1. Erstellen Sie ein Domänenbenutzerkonto für die UNIX-Maschine.
2. Ändern Sie den Hostnamen. Wenn Sie RedHat Linux verwenden, öffnen Sie die Datei `/etc/sysconfig/network` und ändern Sie `HOSTNAME` in die Form `<Name>.<Realm>`. Dadurch kann Active Directory die Serverberechtigungs-nachweise finden.
3. Damit der DNS-Server die UNIX-Maschine finden kann, führen Sie einen der folgenden Schritte aus:
 - Öffnen Sie die Datei `%windows%/system32/drivers/etc/hosts` und fügen Sie die IP/Hostzuordnung hinzu, beispielsweise:
`192.168.1.102 test.modelersso.com test`
 Oder
 - Fügen Sie einen neuen Zoneneintrag für die umgekehrte Adressauflösung hinzu. Auf diese Weise wird auf dem DNS-Server eine IP/Hostzuordnung hinzugefügt.

Wenn der DNS-Eintrag für die UNIX-Maschine nicht korrekt ist, können Sie den Eintrag für die umgekehrte Adressauflösung auf dem DNS-Server manuell hinzufügen.

4. Folgen Sie den Anweisungen zur Konfiguration des IBM SPSS Modeler-Clients, die Sie im Thema „Konfigurieren von Single Sign-on unter Windows“ auf Seite 16 finden.

Hinzufügen und Bearbeiten der IBM SPSS Modeler Server-Verbindung

Serververbindungen können manuell im Kontrollkästchen "Anmelden beim Server" bearbeitet bzw. hinzugefügt werden. Durch Klicken auf "Hinzufügen" können Sie auf ein leeres Dialogfeld vom Typ "Server hinzufügen/bearbeiten" zugreifen, in dem Sie Details zur Serververbindung eingeben können. Durch Auswahl einer bestehenden Verbindung und Klicken auf "Bearbeiten" im Dialogfeld "Anmelden beim Server" wird das Dialogfeld "Server hinzufügen/bearbeiten" mit den Details für die betreffende Verbindung geöffnet, sodass Sie etwaige Änderungen vornehmen können.

Hinweis: Serververbindungen, die über IBM SPSS Collaboration and Deployment Services hinzugefügt wurden, können nicht bearbeitet werden, da Name, Port und sonstige Details in IBM SPSS Collaboration and Deployment Services definiert sind.

So fügen Sie Serververbindungen hinzu:

1. Klicken Sie im Menü "Extras" auf die Option **Anmelden beim Server**. Das Dialogfeld "Anmelden beim Server" wird geöffnet.
 2. Klicken Sie in diesem Dialogfeld auf **Hinzufügen**. Das Dialogfeld "Anmeldung beim Server: Server hinzufügen/bearbeiten" wird angezeigt.
 3. Geben Sie die Details für die Serververbindung ein und klicken Sie auf **OK**, um die Verbindung zu speichern und zum Dialogfeld "Anmeldung beim Server" zurückzukehren.
- **Server**. Geben Sie einen verfügbaren Server an oder wählen Sie einen aus der Liste aus. Der Server-Computer lässt sich anhand eines alphanumerischen Namens (z. B. *meinserver*) oder der dem Server-Computer zugewiesenen IP-Adresse (z. B. 202.123.456.78) identifizieren.
 - **Port**. Geben Sie die Portnummer an, die der Server überwacht. Wenn der Standardwert nicht funktioniert, fragen Sie Ihren Systemadministrator nach der richtigen Portnummer.
 - **Beschreibung**. Geben Sie eine optionale Beschreibung für diese Serververbindung ein.
 - **Verbindung verschlüsseln (mit SSL)**. Legt fest, ob eine SSL-Verbindung (**Secure Sockets Layer**) verwendet werden soll. SSL ist ein weit verbreitetes Protokoll zum Schutz der über ein Netz versendeten Daten. Um diese Funktion verwenden zu können, muss SSL auf dem Server, auf dem sich IBM SPSS Modeler Server befindet, aktiviert sein. Wenden Sie sich gegebenenfalls an den lokalen Administrator, wenn Sie weitere Informationen benötigen.

So bearbeiten Sie Serververbindungen:

1. Klicken Sie im Menü "Extras" auf die Option **Anmelden beim Server**. Das Dialogfeld "Anmelden beim Server" wird geöffnet.
2. Wählen Sie in diesem Dialogfeld die zu bearbeitende Verbindung aus und klicken Sie dann auf **Bearbeiten**. Das Dialogfeld "Anmeldung beim Server: Server hinzufügen/bearbeiten" wird angezeigt.
3. Ändern Sie die Details für die Serververbindung und klicken Sie auf **OK**, um die Änderungen zu speichern und zum Dialogfeld "Anmeldung beim Server" zurückzukehren.

Suchen nach Servern in IBM SPSS Collaboration and Deployment Services

Anstatt eine Serververbindung manuell einzugeben, können Sie einen im Netz verfügbaren Server oder Server-Cluster über Coordinator of Processes auswählen. Diese Funktion ist in IBM SPSS Collaboration and Deployment Services verfügbar. Ein Server-Cluster ist eine Gruppe von Servern, aus denen Coordinator of Processes den Server ermittelt, der am besten für die Beantwortung einer Verarbeitungsanforderung geeignet ist.

Sie können zwar auch manuell Server im Dialogfeld "Anmelden beim Server" hinzufügen, durch die Suche nach verfügbaren Servern können Sie jedoch eine Verbindung zu Servern herstellen, ohne den richtigen Servernamen und die Portnummer kennen zu müssen. Diese Informationen werden automatisch bereitgestellt. Allerdings benötigen Sie auch bei dieser Variante die richtigen Anmeldeinformationen, wie Benutzername, Domäne und Kennwort.

Hinweis: Wenn Sie keinen Zugriff auf die Funktion Coordinator of Processes haben, können Sie dennoch den Namen des Servers, mit dem eine Verbindung hergestellt werden soll, manuell eingeben oder einen zuvor definierten Namen auswählen. Weitere Informationen finden Sie unter dem Thema „Hinzufügen und Bearbeiten der IBM SPSS Modeler Server-Verbindung“ auf Seite 17.

So suchen Sie nach Servern und Clustern:

1. Klicken Sie im Menü "Extras" auf die Option **Anmelden beim Server**. Das Dialogfeld "Anmelden beim Server" wird geöffnet.
2. Klicken Sie in diesem Dialogfeld auf **Suche**, um das Dialogfeld "Nach Servern suchen" zu öffnen. Wenn Sie versuchen, Coordinator of Processes zu durchsuchen, ohne bei IBM SPSS Collaboration and Deployment Services angemeldet zu sein, werden Sie zur Anmeldung aufgefordert.

3. Wählen Sie den Server bzw. Server-Cluster in der Liste aus.
4. Klicken Sie auf **OK**, um das Dialogfeld zu schließen und diese Verbindung der Tabelle im Dialogfeld "Anmelden beim Server" hinzuzufügen.

Daten- und Dateisysteme

Benutzer, die mit IBM SPSS Modeler Server arbeiten, müssen wahrscheinlich auf Datendateien und andere Datenquellen im Netz zugreifen sowie Dateien auf dem Netz speichern. Je nachdem benötigen sie die folgenden Informationen:

- **ODBC-Datenquelle.** Wenn Benutzer auf ODBC-Datenquellen zugreifen müssen, die auf dem Server-Computer definiert sind, benötigen Sie die Namen, Beschreibungen und Anmeldeinformationen (inklusive Datenanmelde-IDs und Kennwörter) für die Datenquellen.
- **Datendateizugriff.** Wenn Benutzer auf dem Server-Computer oder an einem anderen Standort im Netz auf Datendateien zugreifen wollen, müssen Sie die Namen und Speicherorte der Datendateien kennen.
- **Speicherort zum Speichern von Dateien.** Wenn Benutzer während einer Verbindung mit IBM SPSS Modeler Server Daten speichern, kann es vorkommen, dass sie die Dateien auf dem Server-Computer speichern möchten. Der Server ist jedoch häufig schreibgeschützt. In diesem Fall müssen Sie den Benutzern mitteilen, an welchem Speicherort sie die Datendateien speichern sollen. (Normalerweise handelt es sich bei diesem Speicherort um das Basisverzeichnis des Benutzers.)

Benutzerauthentifizierung

IBM SPSS Modeler Server verwendet das Betriebssystem des Server-Computers zum Authentifizieren der Benutzer, die eine Verbindung zum Server herstellen. Wenn ein Benutzer eine Verbindung mit IBM SPSS Modeler Server herstellt, werden alle für den Benutzer ausgeführten Vorgänge im Sicherheitskontext des Benutzers ausgeführt. Der Zugriff auf Datenbanktabellen unterliegt den Benutzer- und/oder Kennwortberechtigungen für die Datenbank.

Windows. Unter Windows kann sich jeder Benutzer anmelden, der im Hostnetz über ein gültiges Konto verfügt. Bei der Standardauthentifizierung benötigen Benutzer Zugriff einschließlich Änderungsrecht auf das Verzeichnis `<Modeler-Server-Installation>\Tmp`. Ohne dieses Recht können sich die Benutzer unter Windows vom Client aus nicht mit der Standardauthentifizierung bei IBM SPSS Modeler Server anmelden.

UNIX. Standardmäßig wird IBM SPSS Modeler Server unter UNIX als Root ausgeführt. Auf diese Weise kann sich jeder Benutzer anmelden, der im Hostnetz über ein gültiges Konto verfügt. Außerdem wird der Zugriff auf die eigenen Dateien und Verzeichnisse eingeschränkt. Sie können jedoch IBM SPSS Modeler Server so konfigurieren, dass er ohne die Rootberechtigung ausgeführt werden kann. In diesem Fall müssen Sie eine private Kennwortdatenbank erstellen, die für die Authentifizierung verwendet wird. Alle IBM SPSS Modeler-Benutzer müssen ein gemeinsames UNIX-Benutzerkonto verwenden (und somit gemeinsamen Zugriff auf Datendateien haben). Weitere Informationen finden Sie im Thema „Konfigurieren als Nicht-Root-Benutzer mit einer privaten Kennwortdatenbank“ auf Seite 64.

Unter Solaris, HP-UX, Linux und AIX-Plattformen verwendet IBM SPSS Modeler Server PAM zur Authentifizierung. Sie können den Servicennamen `modelerserver` zum Konfigurieren der PAM-Module für IBM SPSS Modeler Server verwenden, falls erforderlich.

Dateierstellung

Wenn IBM SPSS Modeler Server auf Daten zugreift und diese verarbeitet, muss das Programm häufig eine temporäre Kopie dieser Daten auf der Festplatte speichern. Die Menge an erforderlichem Festplattenspeicher für die temporären Dateien hängt von der Größe der vom Endbenutzer analysierten Datendatei ab sowie von der jeweils durchgeführten Analyse. Weitere Informationen finden Sie im Thema „Temporäre Festplattenspeicher- und Arbeitsspeicheranforderungen“ auf Seite 10.

UNIX. Die UNIX-Versionen von IBM SPSS Modeler Server verwenden den UNIX-Befehl `umask`, um für die temporären Dateien Dateiberechtigungen festzulegen. Sie können die Standardberechtigungen des Servers überschreiben. Weitere Informationen finden Sie im Thema „Steuern der Berechtigungen für die Dateierstellung“ auf Seite 56.

Unterschiede bei den Ergebnissen

Benutzer, die in beiden Modi Analysen durchführen, erhalten möglicherweise leichte Unterschiede bei den Ergebnissen von IBM SPSS Modeler und IBM SPSS Modeler Server. Diese Diskrepanz tritt normalerweise auf, wenn Datensätze anders sortiert oder gerundet werden.

Sortieren von Datensätzen. Wenn ein Stream Datensätze nicht ausdrücklich sortiert, kann die Reihenfolge der dargestellten Datensätze zwischen lokal und auf dem Server ausgeführten Streams voneinander leicht abweichen. Außerdem kann es zwischen Vorgängen, die in einer Datenbank ausgeführt werden, und Vorgängen, die in IBM SPSS Modeler Server ausgeführt werden, zu Unterschieden in der Reihenfolge kommen. Die Ursache für diese Unterschiede liegt in den verschiedenen Algorithmen, die jedes System zum Implementieren von Funktionen zum erneuten Sortieren der Datensätze verwendet, beispielsweise die Aggregation. Außerdem legt SQL die Reihenfolge, in der Datensätze aus einer Datenbank zurückgegeben werden, ohne ausdrücklichen Sortiervorgang nicht fest.

Unterschiede beim Runden. IBM SPSS Modeler verwendet im lokalen Modus ein anderes internes Format zum Speichern von Gleitkommawerten als IBM SPSS Modeler Server. Aufgrund der Unterschiede beim Runden können die Ergebnisse leicht voneinander abweichen.

Kapitel 4. IBM SPSS Modeler Server-Verwaltung

Dieses Kapitel enthält Informationen zum Starten und Stoppen von IBM SPSS Modeler Server, zum Konfigurieren verschiedener Serveroptionen und zur Interpretation der Protokolldatei. Es beschreibt die Verwendung von IBM SPSS Modeler Administration Console, einer Anwendung zur Konfiguration und Überwachung des Servers. Installationsanweisungen zu dieser Komponente finden Sie in den Installationsanweisungen zu IBM SPSS Modeler Server, die im Lieferumfang des Produkts enthalten sind.

Starten und Stoppen von IBM SPSS Modeler Server

IBM SPSS Modeler Server wird unter Windows als Dienst und unter UNIX als Dämonprozess ausgeführt.

Hinweis zur Planung: Beim Stoppen von IBM SPSS Modeler Server wird die Verbindung der einzelnen Endbenutzer unterbrochen und ihre Sitzung wird beendet. Versuchen Sie also, Serverneustarts in Zeiten geringer Nutzung zu legen. Wenn dies nicht möglich ist, sollten Sie die Benutzer unbedingt vor dem Anhalten des Servers benachrichtigen.

Starten, Stoppen und Statusprüfung unter Windows

Unter Windows steuern Sie IBM SPSS Modeler Server mit dem Dialogfeld "Dienste" in der Windows-Systemsteuerung.

1. **Windows XP.** Öffnen Sie das Windows-Startmenü. Wählen Sie **Einstellungen** und dann **Systemsteuerung**. Doppelklicken Sie auf **Verwaltung** und dann auf **Dienste**.
Windows 2003 oder 2008. Öffnen Sie das Windows-Startmenü. Wählen Sie **Systemsteuerung**, dann **Verwaltung** und anschließend **Dienste** aus.
2. Wählen Sie den Dienst **IBM SPSS Modeler Server <nn.n>** aus. Sie können nun seinen Status überprüfen, ihn starten bzw. stoppen und nach Bedarf die Startparameter bearbeiten.

Standardmäßig ist der Dienst für automatischen Start konfiguriert, d. h., wenn Sie ihn stoppen, wird er automatisch bei einem erneuten Booten des Computers neu gestartet. Wenn der Dienst auf diese Weise gestartet wird, wird er unbeaufsichtigt ausgeführt und der Computer kann abgemeldet werden, ohne dass dies Einfluss auf den Dienst hätte.

Starten, Stoppen und Statusprüfung unter UNIX

Unter UNIX können Sie IBM SPSS Modeler Server starten bzw. stoppen, indem Sie das Script *modelersrv.sh* im Installationsverzeichnis von IBM SPSS Modeler Server ausführen.

1. Wechseln Sie zum IBM SPSS Modeler Server-Installationsverzeichnis. Geben Sie beispielsweise an einer UNIX-Eingabeaufforderung Folgendes ein:
`cd /usr/modelersrv`
Dabei steht *modelersrv* für das Installationsverzeichnis von IBM SPSS Modeler Server.
2. Um den Server zu starten, geben Sie Folgendes an der Eingabeaufforderung ein:
`./modelersrv.sh start`
3. Um den Server zu stoppen, geben Sie Folgendes an der Eingabeaufforderung ein:
`./modelersrv.sh stop`
4. Um den Status von IBM SPSS Modeler Server zu überprüfen, geben Sie an einer UNIX-Eingabeaufforderung Folgendes ein:
`./modelersrv.sh list`

Untersuchen Sie dann die Ausgabe, die den Ausgaben des UNIX-Befehls `ps` ähnelt. Der erste Prozess in der Liste ist der IBM SPSS Modeler Server-Dämonprozess, die restlichen Prozesse sind IBM SPSS Modeler-Sitzungen.

Das IBM SPSS Modeler Server-Installationsprogramm beinhaltet ein Script (*auto.sh*), das das System so konfiguriert, dass der Serverdämon automatisch beim Booten gestartet wird. Wenn Sie dieses Script ausgeführt haben und dann den Server stoppen, wird der Dämon automatisch neu gestartet, wenn der Computer erneut gebootet wird. Weitere Informationen finden Sie im Thema „Automatisches Starten und Stoppen von IBM SPSS Modeler Server“ auf Seite 55.

Grenzwerte für UNIX-Kernel

Sie müssen sicherstellen, dass Kernelgrenzwerte im System für den Betrieb von IBM SPSS Modeler Server ausreichen. Die "ulimits" für Daten, Arbeitsspeicher und Dateien sind besonders wichtig und sollten in der IBM SPSS Modeler Server-Umgebung auf unbegrenzt eingestellt werden. Gehen Sie dazu wie folgt vor:

1. Fügen Sie *modelersrv.sh* die folgenden Befehle hinzu:

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

2. Starten Sie IBM SPSS Modeler Server neu.

Umgang mit nicht reagierenden Serverprozessen (UNIX-Systeme)

IBM SPSS Modeler Server-Prozesse können aus unterschiedlichen Gründen nicht mehr reagieren, z. B. wenn sie einen System- oder ODBC-Treiberaufruf starten, der blockiert wird (Aufruf wird nicht zurückgegeben oder der Vorgang dauert sehr lange). Tritt dies bei UNIX-Prozessen auf, können sie mithilfe des `kill`-Befehls in UNIX bereinigt werden (durch den IBM SPSS Modeler-Client ausgelöste Interrupts oder das Schließen des IBM SPSS Modeler-Clients sind wirkungslos). Ein `kill`-Befehl ist als Alternative zum normalen `stop`-Befehl vorgesehen und ermöglicht dem Administrator die Verwendung von *modelersrv.sh*, um den geeigneten `kill`-Befehl auf einfache Weise auszugeben.

Bei Systemen, auf denen sich nicht verwendbare Serverprozesse ("Geisterprozesse") häufen könnten, wird empfohlen, IBM SPSS Modeler Server in regelmäßigen Intervallen mit der folgenden Befehlsfolge zu stoppen und neu zu starten:

```
cd Modeler_Server-Installationsverzeichnis
./modelersrv.sh stop
./modelersrv.sh kill
```

Die IBM SPSS Modeler-Prozesse, die mithilfe des Befehls `modelersrv.sh kill` beendet werden, hinterlassen temporäre Dateien (aus dem temporären Verzeichnis), die manuell entfernt werden müssen. Temporäre Dateien bleiben unter Umständen auch dann zurück, wenn die Anwendung wegen Ressourcenauslastung abstürzt, sowie bei Benutzerinterrupts, Systemabstürzen oder aus anderen Gründen. Es wird daher empfohlen, IBM SPSS Modeler Server nicht nur regelmäßig neu zu starten, sondern alle verbleibenden Dateien aus dem temporären Verzeichnis von IBM SPSS Modeler zu entfernen.

Sobald alle Serverprozesse geschlossen und die temporären Dateien entfernt wurden, kann IBM SPSS Modeler Server sicher neu gestartet werden.

Verwaltung

IBM SPSS Modeler Server weist eine Reihe von konfigurierbaren Optionen auf, die sein Verhalten steuern. Sie können diese Optionen auf zwei verschiedene Weisen festlegen:

- Verwenden Sie die Anwendung IBM SPSS Modeler Administration Console, die aktuellen IBM SPSS Modeler-Kunden kostenlos zur Verfügung steht. Weitere Informationen finden Sie im Thema „Verwenden von IBM SPSS Modeler Administration Console“.
- Verwenden Sie die Textdatei *options.cfg*, die sich im Verzeichnis *[Serverinstallationspfad]/config* befindet. Weitere Informationen finden Sie im Thema „Verwenden der Datei *options.cfg*“ auf Seite 29.

Es wird empfohlen, IBM SPSS Modeler Administration Console als Verwaltungstool zu installieren und zu verwenden und nicht die Datei *options.cfg* zu bearbeiten. Für die Bearbeitung der Datei ist Zugriff auf das Dateisystem von IBM SPSS Modeler Server erforderlich, mit IBM SPSS Modeler Administration Console dagegen können Sie jede Person mit einem Benutzerkonto zur Anpassung dieser Optionen berechtigen. Außerdem bietet IBM SPSS Modeler Administration Console weitere Informationen zu den Serverprozessen, sodass Sie Nutzung und Leistung überwachen können. Anders als bei der Bearbeitung der Konfigurationsdatei können die meisten Konfigurationsoptionen ohne Neustart von IBM SPSS Modeler Server geändert werden.

Weitere Informationen zur Verwendung von IBM SPSS Modeler Administration Console und der Datei *options.cfg* finden Sie in folgenden Abschnitten.

Verwenden von IBM SPSS Modeler Administration Console

Die Anwendung Modeler Administration Console bietet eine Konsolenbenutzerschnittstelle zur Überwachung und Konfiguration der SPSS Modeler Server-Installationen und steht aktuellen SPSS Modeler Server-Kunden kostenlos zur Verfügung. Die Anwendung kann nur unter Windows installiert werden. Der von ihr verwaltete Server kann jedoch auf einer beliebigen unterstützten Plattform installiert sein.

Viele der über Modeler Administration Console verfügbaren Optionen können auch in der Datei *options.cfg* angegeben werden, die sich im Installationsverzeichnis von SPSS Modeler Server unter */config* befindet. Die Anwendung Modeler Administration Console bietet jedoch eine für die gemeinsame Verwendung freigegebene grafische Benutzerschnittstelle, mit der Sie eine Verbindung mit mehreren Servern herstellen und diese konfigurieren und überwachen können.

Starten von Modeler Administration Console

Zeigen Sie im Startmenü von Windows auf **[Alle] Programme, IBM SPSS Collaboration and Deployment Services, Deployment Manager, IBM SPSS Modeler Administration Console <nn.n>**.

Wenn Sie die Anwendung zum ersten Mal ausführen, sehen Sie leere Serveradministrations- und Eigenschaftsbereiche (außer Sie haben Deployment Manager bereits mit einer schon eingerichteten IBM SPSS Collaboration and Deployment Services-Serververbindung installiert). Nach der Konfiguration von Modeler Administration Console wird im Serveradministrationsbereich auf der linken Seite ein Knoten für jeden SPSS Modeler Server angezeigt, den Sie verwalten können. Im rechten Bereich werden die Konfigurationsoptionen für den ausgewählten Server angezeigt. Sie müssen zuerst für jeden Server, den Sie verwalten möchten, eine Verbindung einrichten.

Konfigurieren des Zugriffs mit Modeler Administration Console

Der Administratorzugriff auf SPSS Modeler Server über Modeler Administration Console wird über die Zeile *administrators* in der Datei *options.cfg* gesteuert. Diese Datei befindet sich im Installationsverzeichnis von SPSS Modeler Server unter */config*. Diese Zeile ist standardmäßig auf Kommentar gesetzt. Sie müssen diese Zeile also bearbeiten, um den Zugriff durch bestimmte Personen zuzulassen, oder *** verwenden, um allen Benutzern Zugriff zu gewähren, wie in den folgenden Beispielen gezeigt:

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

- Die Zeile muss mit *administrators*, beginnen und die Einträge müssen in Anführungszeichen gesetzt werden. Bei den Einträgen wird zwischen Groß- und Kleinschreibung unterschieden.
- Trennen Sie mehrere Benutzer-IDs mit Kommas.

- Verwenden Sie im Falle von Windows-Konten keine Domännennamen.
- Verwenden Sie das Sternchen (*) umsichtig. Mit dieser Einstellung kann sich jede Person mit einem gültigen Benutzerkonto für IBM SPSS Modeler Server (meistens jede beliebige Person im Netz) anmelden und die Konfigurationsoptionen ändern.

SPSS Modeler Server-Verbindungen

Sie müssen für jede SPSS Modeler Server-Instanz in Ihrem Netz, die Sie verwalten möchten, eine Verbindung einrichten. Anschließend müssen Sie sich bei jedem Server anmelden. Die Serververbindung wird für Modeler Administration Console-Sitzungen gespeichert, nicht jedoch die Anmeldeberechtigungsanzeige. Sie müssen sich bei jedem Start von Modeler Administration Console erneut anmelden.

So richten Sie eine Serververbindung ein

1. Stellen Sie sicher, dass der IBM SPSS Modeler Server-Service gestartet wurde.
2. Wählen Sie im Menü "Datei" die Option **Neu** und dann **Verwaltete Serververbindung** aus.
3. Geben Sie in der ersten Seite des Assistenten einen Namen für die Serververbindung ein. Der Name ist zu Ihrer eigenen Verwendung und sollte beschreibend sein, zum Beispiel *Produktionsserver*. Stellen Sie sicher, dass der Typ auf **Verwaltet IBM SPSS Modeler Server** gesetzt ist und klicken Sie dann auf **Weiter**.
4. Auf der zweiten Seite geben Sie den Hostnamen oder die IP-Adresse des Servers ein. Wenn Sie die Standardeinstellung für den Port geändert haben, geben Sie die Portnummer ein. Klicken Sie auf **Fertigstellen**. Die neue Serververbindung wird im Serveradministrationsbereich angezeigt.

Um Verwaltungsaufgaben durchzuführen, müssen Sie sich anmelden.

So melden Sie sich beim Server an

1. Doppelklicken Sie im Serveradministrationsbereich auf den Server, bei dem Sie sich anmelden möchten.
2. Geben Sie im Dialogfeld "Anmelden" Ihre Berechtigungsnachweise ein. (Verwenden Sie Ihr Benutzerkonto für den Serverhost.) Klicken Sie auf **OK**.

Wenn die Anmeldung fehlschlägt und die Nachricht **Administratorrechte auf dem Server nicht verfügbar** erscheint, ist die wahrscheinlichste Ursache, dass der Administratorzugriff nicht korrekt konfiguriert wurde. Weitere Informationen finden Sie im Thema „Konfigurieren des Zugriffs mit Modeler Administration Console“ auf Seite 23.

Wenn die Anmeldung fehlschlägt und die Fehlermeldung **Verbindung zum Server '<Server>' fehlgeschlagen** angezeigt wird, überprüfen Sie, dass die Benutzer-ID und das Kennwort korrekt sind und dass der IBM SPSS Modeler Server-Service ausgeführt wird. Rufen Sie z. B. unter Windows "Systemsteuerung > Verwaltung > Dienste" auf und überprüfen Sie den Eintrag für IBM SPSS Modeler Server. Wenn die Statusspalte nicht **Gestartet** anzeigt, wählen Sie diese Zeile auf dem Bildschirm aus und klicken Sie auf **Starten**. Versuchen Sie anschließend erneut, sich anzumelden.

Nach der Anmeldung bei IBM SPSS Modeler Server werden unter dem Servernamen die beiden Optionen Konfiguration und Überwachung angezeigt. Doppelklicken Sie auf eine dieser Optionen.

SPSS Modeler Server-Konfiguration

Im Konfigurationsbereich werden Konfigurationsoptionen für SPSS Modeler Server angezeigt. In diesem Bereich können Sie die Optionen wunschgemäß ändern. Klicken Sie in der Symbolleiste auf **Speichern**, um die Änderungen zu speichern. (Beachten Sie, dass eine Änderung der Portnummer erst nach dem Neustart des Servers wirksam wird.)

Die Optionen werden in den nachfolgenden Abschnitten beschrieben und die entsprechende Zeile in der Datei *options.cfg* wird für jede Option in Klammern angegeben. Optionen, die nur in *options.cfg* sichtbar sind, werden am Ende dieses Abschnitts beschrieben.

Verbindungen/Sitzungen

Maximale Anzahl an Verbindungen. (max_sessions) Maximale Anzahl an Benutzern, die gleichzeitig verbunden sein können. Der Wert -1 bedeutet, dass keinerlei Beschränkung vorliegt.

Portnummer. (port_number) Die Portnummer, die SPSS Modeler Server überwacht. Ändern Sie die Portnummer, wenn der Standardport bereits von einer anderen Anwendung verwendet wird. Endbenutzer müssen die Portnummer kennen, um mit SPSS Modeler Server arbeiten zu können.

Analytic Server-Verbindung

Analytic Server-URL. (as_url) Die URL von Analytic Server, einschließlich der IP-Adresse des Servers, des Ports und des Nutzers (Tenants), zu dem die SPSS Modeler Server-Installation gehört.

Eingabeaufforderung für Kennwort. (as_prompt_for_password) Geben Sie N an, wenn SPSS Modeler Server mit demselben Authentifizierungssystem für Benutzer und Kennwörter konfiguriert wird, das auch für Analytic Server verwendet wird. Geben Sie andernfalls Y an.

Datendateizugriff

Zugriff ausschließlich auf Datendateipfad beschränken. (data_files_restricted) Ist dieser Wert auf *Ja* eingestellt, wird der Zugriff auf Datendateien im standardmäßigen Datenverzeichnis und den im **Datendateipfad** aufgelisteten Dateien eingeschränkt.

Datendateipfad. (data_file_path) Eine Liste zusätzlicher Verzeichnisse, für die Clients Lese- und Schreibrechte besitzen. Diese Option wird ignoriert, außer die Option **Zugriff auf Datendateipfad beschränken** ist aktiviert. Beachten Sie, dass bei allen Pfadangaben normale Schrägstriche (/) zu verwenden sind. Trennen Sie mehrere Verzeichnisse durch Semikolons (z. B. *[Serverinstallationspfad]/data;c:/data;c:/temp*).

Zugriff ausschließlich auf Programmdateipfad beschränken. (program_files_restricted) Ist dieser Wert auf *Ja* eingestellt, wird der Zugriff auf Programmdateien im Standardverzeichnis *bin* und den im **Programmdateipfad** aufgelisteten Dateien eingeschränkt.

Programmdateipfad. (program_file_path) Eine Liste zusätzlicher Verzeichnisse, in denen Clients Programme ausführen dürfen. Diese Option wird ignoriert, außer die Option **Zugriff ausschließlich auf Programmdateipfad beschränken** ist aktiviert. Beachten Sie, dass bei allen Pfadangaben normale Schrägstriche (/) zu verwenden sind. Trennen Sie mehrere Verzeichnisse durch Semikolons.

Maximale Dateigröße (MB). (max_file_size) Maximale Größe von temporären und exportierten Datendateien, die während der Streamausführung erstellt wurden (gilt nicht für SAS- und SPSS Statistics-Datendateien). Der Wert -1 bedeutet, dass keinerlei Beschränkung vorliegt.

Temporäres Verzeichnis. (temp_directory) Das zum Speichern temporärer Datendateien (Cachedateien) verwendete Verzeichnis. Idealerweise sollte sich dieses Verzeichnis auf einem separaten Hochgeschwindigkeitslaufwerk oder Controller befinden, da die Zugriffszeit auf dieses Verzeichnis erhebliche Auswirkungen auf die Leistung haben kann. Es können mehrere temporäre Verzeichnisse (durch Komma getrennt) angegeben werden. Diese sollten sich auf unterschiedlichen Datenträgern befinden; das erste Verzeichnis wird am häufigsten verwendet, die zusätzlichen Verzeichnisse dienen zur Speicherung temporärer Arbeitsdateien, wenn bestimmte Datenaufbereitungsvorgänge (beispielsweise die Sortierung) bei der Ausführung Parallelität verwenden. Dadurch, dass die einzelnen Ausführungsthreads gesonderte Datenträger für die temporäre Speicherung verwenden können, ist eine Leistungssteigerung möglich. Verwenden Sie bei allen Pfadangaben normale Schrägstriche (/). *Hinweis:* Beim Starten von SPSS Modeler Server werden in diesem Verzeichnis temporäre Dateien generiert. Stellen Sie sicher, dass Sie über die

notwendigen Zugriffsrechte für dieses Verzeichnis verfügen (beispielsweise wenn es sich beim temporären Verzeichnis um einen freigegebenen Netzordner handelt), andernfalls kann SPSS Modeler Server nicht gestartet werden.

Pfad zur ausführbaren Python-Datei. (`python_exe_path`) Vollständiger Pfad zur ausführbaren Python-Datei einschließlich Name der exe-Datei. Je nachdem, wo Python installiert ist, kann es erforderlich sein, dass für **Zugriff ausschließlich auf Programmdateipfad beschränken** die Option *Nein* angegeben sein muss.

Leistung und Optimierung

Streamneuerstellung. (`stream_rewriting_enabled`) Ermöglicht dem Server die Optimierung von Streams durch Neuschreiben. Der Server kann beispielsweise die Datenreduktionsvorgänge näher an den Quellknoten bringen, um die Größe des Datasets so früh wie möglich zu verringern. Die Inaktivierung dieser Option wird normalerweise nur empfohlen, wenn durch die Optimierung Fehler oder andere unerwartete Ergebnisse verursacht werden. Diese Einstellung überschreibt die entsprechende Optimierungseinstellung am Client.

Parallelismus. (`max_parallelism`) Gibt die Anzahl an parallelen Arbeitsthreads an, die SPSS Modeler beim Ausführen eines Streams verwenden darf. Wenn Sie diesen Wert auf 0 oder eine negative Zahl setzen, wird IBM SPSS Modeler angewiesen, die Anzahl an Threads der Anzahl an verfügbaren Prozessoren auf dem Computer anzupassen. Der Standardwert für diese Option ist -1. Um die parallele Verarbeitung (für Computer mit mehreren Prozessoren) zu inaktivieren, setzen Sie diese Option auf den Wert 1. Um eine eingeschränkte parallele Verarbeitung zu gestatten, setzen Sie den Wert auf eine Zahl, die geringer als die auf Ihrem Computer verfügbare Anzahl an Prozessoren ist. Beachten Sie, dass ein Hyperthreaded- oder Dual-Core-Prozessor als zwei Prozessoren betrachtet wird.

Puffergröße (Byte). (`io_buffer_size`) Die vom Server zum Client übertragenen Datendateien werden durch einen Puffer mit dieser Größe geleitet.

Cachekomprimierung. (`cache_compression`) Ein ganzzahliger Wert von 0 bis 9, der die Komprimierung des Cache und anderer Dateien im temporären Verzeichnis des Servers steuert. Durch Komprimierung reduziert sich die verwendete Speichermenge, was bei begrenzter Speicherkapazität von Bedeutung ist. Außerdem wird die Leistungsfähigkeit erhöht, da weniger Datenträgeraktivität zum Erstellen und Lesen des Cache erforderlich ist. Durch die Komprimierung erhöht sich die Prozessorzeit; dies wird jedoch fast immer durch die verringerte Datenträgerzugriffszeit ausgeglichen. Beachten Sie, dass eine Komprimierung nur bei bestimmten Caches möglich ist, nämlich bei solchen, auf die der Zugriff sequenziell erfolgt. Diese Option steht nicht bei Caches mit Zufallszugriff zur Verfügung, beispielsweise denjenigen, die von den Netztrainingsalgorithmen verwendet werden. Mit dem Wert 0 wird die Komprimierung vollständig inaktiviert. Die Werte von 1 aufwärts bieten einen steigenden Komprimierungsgrad, jedoch jeweils mit entsprechend erhöhter Zugriffszeit. Der Standardwert ist 1; höhere Werte sind selten erforderlich, es sei denn, der Datenträgerspeicher ist besonders knapp.

Multiplikator für Arbeitsspeicherverwendung. (`memory_usage`) Legt fest, welcher Anteil an physischem Speicher für die Sortierung und andere speicherinterne Cachevorgänge zugeordnet wird. Die Standardeinstellung ist 100. Dies entspricht etwa 10 % des physischen Speichers. Wenn freier Arbeitsspeicher verfügbar ist, können Sie diesen Wert erhöhen, um die Sortierleistung zu verbessern. Sie sollten jedoch bei der Erhöhung vorsichtig vorgehen, um übermäßig große Auslagerungsdateien zu vermeiden.

Obergrenze für Modellierungsarbeitsspeicher (Prozent). (`modelling_memory_limit_percentage`) Steuert, welcher Anteil an physischem Speicher für das Training von Modellen für neuronale Netze sowie Kohonen- und K-Means-Modellen zugeordnet wird. Der Standardwert ist 25 %. Wenn freier Arbeitsspeicher verfügbar ist, können Sie diesen Wert erhöhen, um die Trainingsleistung zu verbessern. Sie sollten jedoch bei der Erhöhung vorsichtig vorgehen, damit Sie den Wert nicht so hoch setzen, dass die Auslagerungsdatei übermäßig groß wird, wenn Daten auf den Datenträger geschrieben werden müssen.

Umgehung von Modellierungsarbeitsspeicher erlauben. (`allow_modelling_memory_override`) Aktiviert bzw. inaktiviert die Option **Auf Geschwindigkeit optimiert** in bestimmten Modellierungsknoten. Diese Option ist standardmäßig aktiviert. Mit dieser Option kann der Modellierungsalgorithmus den gesamten Arbeitsspeicher beanspruchen (die Option für die Prozentgrenze wird übergangen). Diese Option sollte inaktiviert werden, wenn Arbeitsspeicherressourcen auf dem Server-Computer für die gemeinsame Verwendung freigegeben sein müssen.

Maximale und minimale Nummer für Server-Port. (`max_server_port` und `min_server_port`) Gibt den Bereich der Portnummern an, die für die zusätzlichen Socketverbindungen zwischen Client und Server verwendet werden können, die für interaktive Modelle und die Streamausführung erforderlich sind. Dafür muss der Server einen anderen Port überwachen; wenn der Bereich nicht eingeschränkt wird, kann dies zu Problemen für Benutzer führen, die Systeme mit Firewalls verwenden. Der Standardwert für beide Optionen lautet -1, was bedeutet, dass es keine Einschränkungen gibt. Um also beispielsweise festzulegen, dass der Server Port 8000 oder höher überwachen soll, setzen Sie `min_server_port` auf 8000 und `max_server_port` auf 1.

Beachten Sie, dass Sie zusätzliche Ports über den Hauptserverport öffnen müssen, um einen Stream zu öffnen oder auszuführen, und entsprechend mehr Ports, wenn Sie mehrere Streams gleichzeitig öffnen oder ausführen möchten. Dies dient zur Erfassung des Feedbacks aus der Streamausführung.

Standardmäßig verwendet IBM SPSS Modeler jeden beliebigen offenen Port, der verfügbar ist. Wenn kein offener Port gefunden wird (z. B. wenn alle durch eine Firewall geschlossen sind) wird bei der Ausführung des Streams eine Fehlernachricht angezeigt. Zur Konfiguration des Portbereichs benötigt IBM SPSS Modeler einen verfügbaren offenen Port (zusätzliche zum Hauptport des Servers) pro gleichzeitigem Stream plus einen zusätzlichen Port für jeden angeschlossenen Client, der über eine ODBC-Verbindung verfügt. Um mehr Streams gleichzeitig auszuführen, können Sie den Bereich der Portnummern erweitern und entsprechend mehr Ports in der Firewall öffnen.

Hinweis: Wenn Sie diese Parameter ändern, müssen Sie SPSS Modeler Server neu starten, damit die Änderungen wirksam werden.

Array-Abrofoptimierung. (`sql_row_array_size`) Legt fest, wie SPSS Modeler Server Daten von der ODBC-Datenquelle abrufen. Der Standardwert ist 1. Dabei wird jeweils eine einzelne Zeile abgerufen. Bei einer Erhöhung dieses Werts liest der Server die Informationen in größeren Einheiten, indem er jeweils die angegebene Anzahl an Zeilen in ein Array abrufen. Bei einigen Kombinationen aus Betriebssystem/Datenbank kann dies zu Verbesserungen bei der Leistungsfähigkeit von SELECT-Anweisungen führen.

SQL

Maximale SQL-Zeichenfolgenlänge. (`max_sql_string_length`) Bei Zeichenfolgen, die mit SQL aus der Datenbank importiert werden, ist dies die maximale Anzahl an Zeichen, die garantiert erfolgreich weitergeleitet werden. Je nach Betriebssystem werden Zeichenfolgewerte, die länger als diese Vorgabe sind, möglicherweise ohne Warnung rechts abgeschnitten. Der zulässige Bereich liegt zwischen 1 und 65.535 Zeichen. Diese Eigenschaft wird auch auf den Datenbankexportknoten angewendet.

Automatische SQL-Generierung. (`sql_generation_enabled`) Ermöglicht automatische SQL-Generierung für Streams, wodurch die Leistung erheblich verbessert werden kann. Diese Option ist standardmäßig aktiviert. Die Inaktivierung dieser Option wird nur empfohlen, wenn die Datenbank keine Anfragen unterstützen kann, die von SPSS Modeler Server übermittelt wurden. Beachten Sie, dass diese Einstellung die entsprechende Optimierungseinstellung am Client überschreibt; beachten Sie außerdem, dass die SQL-Generierung für Scoring-Zwecke unabhängig von dieser Einstellung für jeden Modellierungsknoten gesondert aktiviert werden muss.

Standardlänge für SQL-Zeichenfolgen. (`default_sql_string_length`). Gibt die Standardbreite der Zeichenfolgenspalten an, die in den Datenbankcachetabellen erstellt werden. Zeichenfolgenfelder in Datenbankcachetabellen werden mit einer Standardbreite von 255 erstellt, wenn keine Informationen von einem weiter oben liegenden Typ zur Verfügung stehen. Wenn Sie in Ihren Daten breitere Werte haben, instanzii-

ieren Sie entweder einen weiter oben liegenden Typknoten mit diesen Werten, oder setzen Sie diesen Parameter auf einen Wert, der groß für diese Zeichenfolgewerte ist.

Protokollierung unterstützter SQL-Typen (`supported_sql_type_logging`). Nur zu Debugging-Zwecken; ändern Sie diesen Wert nur, wenn Sie dazu von einem IBM Technical Support-Mitarbeiter angewiesen werden.

Datenbank für benutzerdefinierte Funktionen aktivieren. (`db_udf_enabled`). Wenn diese Option auf Y (Standard) gesetzt ist, generiert die Option für die SQL-Generierung kein reines SPSS Modeler-SQL, sondern UDF-SQL (UDF - User-Defined Function; benutzerdefinierte Funktion). UDF-SQL ist reinem SQL normalerweise überlegen.

SSL

SSL aktivieren. (`ssl_enabled`) Aktiviert die SSL-Verschlüsselung für Verbindungen zwischen SPSS Modeler und SPSS Modeler Server.

Zertifikatsdatei. (`ssl_certificate_file`) Die Datei mit dem öffentlichen SSL-Zertifikat, das beim Start des Servers geladen werden soll.

Datei für privaten Schlüssel. (`ssl_private_key_file`) Die Datei mit dem privaten SSL-Schlüssel, der beim Start des Servers geladen werden soll.

Kennwort für privaten Schlüssel. (`ssl_private_key_password`) Kennwortschlüssel für das angegebene Zertifikat. Dieser Schlüssel muss von mindestens einer der in der konfigurierten Zertifikatsdatei aufgeführten Zertifizierungsstellen signiert worden sein.

Konfiguration von Coordinator of Processes

Host. (`cop_host`) Der Hostname bzw. die IP-Adresse des Coordinator of Processes-Diensts. Der Standardname "spssc" ist ein frei wählbarer Name, den Administratoren als Alias für den IBM SPSS Collaboration and Deployment Services-HOST in DNS hinzufügen können.

Portnummer. (`cop_port_number`) Die Portnummer des Coordinator of Processes-Diensts. Der Standardport, 8080, ist der IBM SPSS Collaboration and Deployment Services-Standardport.

Kontextstammverzeichnis. (`cop_context_root`) Die URL des Coordinator of Processes-Dienst.

Anmeldename. (`cop_user_name`) Der Benutzername für die Authentifizierung beim Coordinator of Processes-Dienst. Dies ist ein IBM SPSS Collaboration and Deployment Services-Anmeldename und kann somit ein Präfix für den Sicherheitsprovider enthalten (z. B.: ad/jsmith).

Kennwort. (`cop_password`) Das Kennwort für die Authentifizierung beim Coordinator of Processes-Dienst.

Aktiviert. (`cop_enabled`) Legt fest, ob der Server versuchen sollte, eine Registrierung beim Coordinator of Processes durchzuführen. Standardmäßig erfolgt *keine* Registrierung, da der Administrator auswählen sollte, welche Dienste über den Coordinator of Processes angekündigt werden.

Serververname. (`cop_service_name`) Der Name dieser SPSS Modeler Server-Instanz. Standardmäßig wird der Hostname verwendet.

Beschreibung. (`cop_service_description`) Eine Beschreibung dieser Instanz.

Aktualisierungsintervall (Min.). (`cop_update_interval`) Die Anzahl der Minuten zwischen Keep-Alive-Nachrichten; der Standardwert ist 2.

Gewichtung. (`cop_service_weight`) Die Gewichtung der betreffenden Instanz als ganze Zahl zwischen 1 und 10. Ein höherer Wert zieht mehr Verbindungen an. Der Standardwert lautet 1.

Service-Host. (cop_service_host) Der vollqualifizierte Hostname des IBM SPSS Modeler Server-Hosts. Der Standardhostname wird automatisch abgeleitet. Bei mehrfach vernetzten Hosts kann der Administrator den Standardwert überschreiben.

In options.cfg sichtbare Optionen

Administratoren. Geben Sie die Benutzernamen der Benutzer an, denen Sie Administratorzugriff gewähren möchten. Weitere Informationen finden Sie im Thema „Konfigurieren des Zugriffs mit Modeler Administration Console“ auf Seite 23.

allow_config_custom_overrides. Ändern Sie diesen Wert nur, wenn Sie dazu von einem IBM Technical Support-Mitarbeiter angewiesen werden.

fips_encryption. Ermöglicht die mit FIPS kompatible Verschlüsselung. Der Standardwert ist N.

max_transfer_size. Nur für systeminterne Verwendung, nicht bearbeiten.

shell. (nur bei UNIX-Servern) Überschreibt die Standardeinstellung für die UNIX-Shell, beispielsweise shell, "/usr/bin/ksh". Standardmäßig verwendet IBM SPSS Modeler die Shell, die in dem Benutzerprofil des Benutzers definiert wurde, der die Verbindung zu IBM SPSS Modeler Server herstellt.

start_process_as_login_user. Wählen Sie hier die Einstellung Y, wenn Sie Single Sign-on verwenden und mit einem Server für IBM SPSS Collaboration and Deployment Services verbunden sind, der nicht dieselbe Authentifizierungsmethode wie der IBM SPSS Modeler Server-Host verwendet. Setzen Sie diese Option auch auf Y, wenn Sie SPSS Modeler Server mit einer privaten Kennwortdatenbank ausführen.

time_encode_execution_log. Setzen Sie diese Option auf Y, wenn die Datei *message.log* (im Ordner *log* der SPSS Modeler Server-Installation) die einzelnen Ausführungszeiten für jeden Knoten im Stream enthalten soll.

use_bigint_for_count. Wenn die Anzahl der zu zählenden Datensätze höher ist als eine normale ganze Zahl ($2^{32}-1$) angeben kann, setzen Sie diese Option auf Y. Wenn diese Option auf Y gesetzt ist und ein Stream entweder mit DB2 oder SQL Server verbunden ist, wird COUNT_BIG() verwendet, wenn eine Datensatzanzahl erforderlich ist (z. B. im vom Aggregatknoten generierten Feld **Datensatzanzahl**).

SPSS Modeler Server-Überwachung

Im Überwachungsbereich von Modeler Administration Console wird eine Momentaufnahme aller aktuell auf dem SPSS Modeler Server-Computer laufenden Prozesse angezeigt, ähnlich wie im Windows Task-Manager. Sie öffnen den Überwachungsbereich, indem Sie im Serveradministratorsbereich auf den Überwachungsknoten unter dem gewünschten Server doppelklicken. Der Bereich wird dann mit einer aktuellen Momentaufnahme der Serverdaten gefüllt. Die Daten werden im angezeigten Intervall aktualisiert (standardmäßig 1 Minute). Um die Daten von Hand zu aktualisieren, klicken Sie auf **Aktualisieren**. Wenn Sie in dieser Liste nur SPSS Modeler Server-Prozesse anzeigen möchten, klicken Sie auf **Nicht-SPSS Modeler -Prozesse herausfiltern**.

Verwenden der Datei options.cfg

Die Datei *options.cfg* befindet sich im Verzeichnis *[Serverinstallationspfad]/config*. Die einzelnen Einstellungen werden als durch Kommata getrenntes Namen-Werte-Paar angegeben, wobei der **Name** den Namen der Option bezeichnet und der **Wert** den Wert für die Option. Rauten (#) zeigen Kommentare an.

Hinweis: Die meisten Konfigurationsoptionen können anstatt mit dieser Konfigurationsdatei mithilfe von IBM SPSS Modeler Administration Console geändert werden, es gibt jedoch einige Ausnahmen. Weitere Informationen finden Sie im Thema „In options.cfg sichtbare Optionen“.

Durch die Verwendung von IBM SPSS Modeler Administration Console können Sie Serverneustarts für alle Optionen mit Ausnahme des Serverports vermeiden. Weitere Informationen finden Sie im Thema „Verwenden von IBM SPSS Modeler Administration Console“ auf Seite 23.

Konfigurationsoptionen, die zur Standarddatei hinzugefügt werden können

Datenbankinternes Caching wird standardmäßig mit IBM SPSS Modeler Server aktiviert. Sie können diese Funktion inaktivieren, indem Sie folgende Zeile in die Datei *options.cfg* einfügen:

```
enable_database_caching, N
```

Dadurch werden temporäre Dateien auf dem Server und nicht in der Datenbank erstellt.

So können Sie die Konfigurationsoptionen für IBM SPSS Modeler Server anzeigen oder ändern:

1. Öffnen Sie die Datei *options.cfg* mit einem Texteditor.
2. Suchen Sie die relevanten Optionen. Eine vollständige Liste der Optionen finden Sie unter „SPSS Modeler Server-Konfiguration“ auf Seite 24.
3. Bearbeiten Sie die Werte nach Bedarf. Beachten Sie, dass bei allen Pfadnamenwerten ein normaler Schrägstrich (/) und kein umgekehrter Schrägstrich (\) als Trennzeichen für den Pfadnamen verwendet werden muss.
4. Speichern Sie die Datei.
5. Stoppen Sie IBM SPSS Modeler Server und starten Sie das Programm erneut, um die Änderungen zu übernehmen. Weitere Informationen finden Sie im Thema „Starten und Stoppen von IBM SPSS Modeler Server“ auf Seite 21.

Schließen nicht verwendeter Datenbankverbindungen

Standardmäßig wird in IBM SPSS Modeler mindestens eine Verbindung in einer Datenbank zwischengespeichert, sobald diese Verbindung hergestellt wurde. Die Datenbanksitzung bleibt auch dann geöffnet, wenn Streams, die für den Datenbankzugriff erforderlich sind, nicht ausgeführt werden.

Durch die Zwischenspeicherung von Datenbankverbindungen kann die Ausführungsdauer verbessert werden, da IBM SPSS Modeler nicht bei jeder Ausführung eines Streams eine neue Verbindung zur Datenbank aufbauen muss. In einigen Umgebungen ist es jedoch wichtig, dass Anwendungen Datenbankressourcen so schnell wie möglich freigeben. Wenn zu viele IBM SPSS Modeler-Sitzungen nicht mehr verwendete Datenbankverbindungen aufrechterhalten, kann dies zu einer Auslastung der Datenbankressourcen führen.

Sie können dies vermeiden, indem Sie die IBM SPSS Modeler-Option *cache_connection* in einer benutzerdefinierten Datenbankkonfigurationsdatei inaktivieren. Dadurch reagiert IBM SPSS Modeler auch stabiler auf Fehler in der Datenbankverbindung (wie etwa Zeitlimitüberschreitungen), die auftreten, wenn Verbindungen über einen längeren Zeitraum von einer IBM SPSS Modeler-Sitzung verwendet werden.

So schließen Sie nicht verwendete Datenbankverbindungen:

1. Suchen Sie das Verzeichnis *[Serverinstallationspfad]/config*.
2. Fügen Sie folgende Datei hinzu (oder öffnen Sie sie, wenn sie bereits existiert):
odbc-custom-properties.cfg
3. Fügen Sie folgende Zeile in die Datei ein:
cache_connection, N
4. Speichern und schließen Sie die Datei.
5. Starten Sie IBM SPSS Modeler Server neu, um die Änderungen zu übernehmen.

Verwenden von SSL zur sicheren Datenübertragung

Secure Sockets Layer (SSL) ist ein Protokoll für die Verschlüsselung von Daten, die zwischen zwei Computern übertragen werden. SSL sorgt dafür, dass die Kommunikation zwischen den Computern sicher ist. SSL kann die Authentifizierung von Benutzername/Kennwort sowie den Inhalt eines Austauschs zwischen einem Server und einem Client verschlüsseln.

Funktionsweise von SSL

SSL beruht auf dem öffentlichen und privaten Schlüssel des Servers sowie einem Zertifikat für den öffentlichen Schlüssel, das die Identität des Servers mit seinem öffentlichen Schlüssel verbindet.

1. Wenn ein Client eine Verbindung zu einem Server aufbaut, authentifiziert der Client den Server mit dem Zertifikat für den öffentlichen Schlüssel.
2. Der Client generiert dann eine Zufallszahl, verschlüsselt die Zahl mit dem öffentlichen Schlüssel des Servers und sendet die verschlüsselte Nachricht zurück an den Server.
3. Der Server entschlüsselt die Zufallszahl mit seinem privaten Schlüssel.
4. Aus der Zufallszahl generieren Server und Client die Sitzungsschlüssel, die zur Verschlüsselung und Entschlüsselung nachfolgender Informationen verwendet werden.

Das Zertifikat für den öffentlichen Schlüssel ist in der Regel von einer Zertifizierungsstelle signiert. Zertifizierungsstellen wie VeriSign und Thawte sind Organisationen, die Sicherheitsdaten, die sich in den Zertifikaten für öffentliche Schlüssel befinden, herausgeben, authentifizieren und verwalten. Im Wesentlichen bestätigt die Zertifizierungsstelle die Identität des Servers. Die Zertifizierungsstelle berechnet gewöhnlich eine Gebühr für ein Zertifikat, jedoch können auch selbstsignierte Zertifikate generiert werden.

Schützen der Client/Server- und Server/Server-Kommunikation durch SSL

Hauptschritte beim Schützen der Client/Server- und Server/Client-Kommunikation durch SSL:

1. Installieren Sie OpenSSL auf dem Server-Computer.
2. Beziehen und installieren Sie das SSL-Zertifikat und die Schlüssel.
3. Aktivieren und konfigurieren Sie SSL in der Server-Verwaltungsanwendung (IBM SPSS Collaboration and Deployment Services Deployment Manager).
4. Installieren Sie ggf. Verschlüsselung mit unbegrenzter Stärke auf den Client-Computern.
5. Fügen Sie das Zertifikat zum Client-Keystore hinzu.
6. Weisen Sie Benutzer an, bei der Verbindung zum Server SSL zu aktivieren.

Anmerkung: Gelegentlich fungiert ein Serverprodukt als Client. Ein Beispiel ist ein IBM SPSS Statistics-Server, der eine Verbindung zum IBM SPSS Collaboration and Deployment Services Repository aufbaut. In diesem Fall ist IBM SPSS Statistics-Server der *Client*.

OpenSSL installieren

Wenn OpenSSL noch nicht auf dem Server verfügbar ist, müssen Sie es installieren.

1. Laden Sie OpenSSL von <http://www.openssl.org/> herunter. Verwenden Sie unbedingt die geeignete OpenSSL-Version für die Serverversion:

Tabelle 1. OpenSSL-Versionen.

Serverprodukt	Kompatible OpenSSL-Version
IBM SPSS Statistics 20-22	1.0.0
IBM SPSS Statistics 17-19 (nicht Linux® on System z®) IBM SPSS Modeler 16 (nicht Linux® on System z®)	0.9.8 und Unterversionen (0.9.8a, 0.9.8b usw.)

Tabelle 1. OpenSSL-Versionen (Forts.).

Serverprodukt	Kompatible OpenSSL-Version
IBM SPSS Statistics 19 (Linux® on System z®)	1.0.0
IBM SPSS Modeler 16 (Linux® on System z®)	

Hinweis: Wenn IBM SPSS Modeler Server for UNIX für SSL konfiguriert ist, versucht das Programm, eine Verbindung mit zwei Bibliotheken mit Namen wie den folgenden herzustellen (.so kann von der Plattform abhängig sein):

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

Bei der Verwendung einer Unterversion von SSL (0.9.8a, 0.9.8b usw.) müssen Sie für diese Bibliotheken symbolische Verbindungen mit den Versionen *xxx.0.9.8* erstellen, nach denen IBM SPSS Modeler Server sucht.

2. Installieren und konfigurieren Sie die Software gemäß den Anweisungen. Wenn Sie OpenSSL selbst erstellen, beachten Sie die folgenden Richtlinien:

Windows. OpenSSL muss mit DLLs (die standardmäßig mehrere Threads verwenden) erstellt werden.

UNIX. OpenSSL muss mehrere Threads (was nicht immer standardmäßig der Fall ist) und gemeinsam verwendete Bibliotheken unterstützen.

3. Stellen Sie sicher, dass das System die SSL-Bibliotheken finden kann:

Windows. Stellen Sie sicher, dass sich die OpenSSL-Module im Systempfad befinden.

UNIX. Hängen Sie die SSL-Bibliotheken an das Ende des in *modelersrv.sh* definierten Pfads zum Laden der Bibliothek an. Fügen Sie unter Solaris beispielsweise die folgende Zeile zum Server-Start-Script hinzu:

```
LD_LIBRARY_PATH_64=$LD_LIBRARY_PATH_64:/usr/local/openssl1; export LD_LIBRARY_PATH_64
```

Beziehen und Installieren von SSL-Zertifikat und -Schlüssel

1. Beziehen Sie ein SSL-Zertifikat und eine Schlüsseldatei. Dazu haben Sie zwei Möglichkeiten.
 - Erwerben Sie sie von einer öffentlichen Zertifizierungsstelle (z. B. Verisign oder Thawte). Die öffentliche Zertifizierungsstelle signiert das Zertifikat, um den Server zu verifizieren, der es verwendet.
 - Generieren Sie Schlüssel und Zertifikatdateien mithilfe einer internen Zertifizierungsstelle für selbstsignierte Zertifikate. OpenSSL bietet ein Zertifikatverwaltungstool für diesen Zweck. Alternativ können Sie im Internet nach Anweisungen zur Erstellung von selbstsignierten SSL-Zertifikaten suchen.
2. Kopieren Sie Zertifikat- und Schlüsseldateien in ein lokales Verzeichnis oder in lokale Verzeichnisse auf dem Server. Der öffentliche und der private Schlüssel können in separaten Verzeichnissen gespeichert werden. Sie können auch in einer einzelnen Datei gespeichert werden. Stellen Sie sicher, dass Zugriffsberechtigungen ein zufälliges Anzeigen des Verzeichnisses verhindern, das den privaten Schlüssel enthält.

Aktivieren und Konfigurieren von SSL im IBM SPSS Collaboration and Deployment Services Deployment Manager

1. Starten Sie die Serververwaltungsanwendung (IBM SPSS Collaboration and Deployment Services Deployment Manager) und stellen Sie eine Verbindung zum Server her.
2. Stellen Sie auf der Konfigurationsseite die Option **Secure Sockets Layer** auf Ja ein.
3. Geben Sie in **SSL-Datei - öffentlicher Schlüssel** den vollständigen Pfad zur Datei mit dem öffentlichen Schlüssel an.
4. Geben Sie in **SSL-Datei - privater Schlüssel** den vollständigen Pfad zur Datei mit dem privaten Schlüssel an.

Hinweis: Wenn öffentlicher und privater Schlüssel in einer Datei gespeichert werden, geben Sie in **SSL-Datei - öffentlicher Schlüssel** und **SSL-Datei - privater Schlüssel** dieselbe Datei an.

5. Wählen Sie die folgenden Befehle aus den Menüs aus:

Datei > Speichern

6. Starten Sie Serverdienst oder -dämon neu. Beim Neustart werden Sie zur Eingabe des SSL-Kennworts aufgefordert. Unter Windows können Sie **Kennwort speichern** wählen, um das Kennwort sicher zu speichern. Damit ist es nicht mehr nötig, bei jedem Serverstart das Kennwort einzugeben.

Einrichten von SSL in IBM SPSS Modeler Server

Wenn zum Zugriff auf die private SSL-Schlüsseldatei ein Kennwort erforderlich ist, hängt die Art der Bereitstellung dieses Kennworts vom Betriebssystem ab:

Windows-Systeme

Bei Windows-Systemen können Sie das Kennwort auf folgende Weisen angeben:

- Geben Sie es in der Datei *options.cfg* mithilfe der Option **ssl_private_key_password** an.
- Geben Sie es mithilfe eines Eingabeaufforderungsfensters ein. Wenn Sie die Option **ssl_private_key_password** aus *options.cfg* weglassen, wird das Eingabeaufforderungsfenster beim Start angezeigt.

UNIX-Systeme

Bei UNIX-Systemen können Sie das Kennwort auf folgende Weisen angeben:

- Geben Sie es in der Datei *options.cfg* mithilfe der Option **ssl_private_key_password** an.
- Fügen Sie der Datei *modelersrv.sh* einen Schalter hinzu, um das Kennwort beim Start anzufordern.

Wenn Sie die zweite Option verwenden, suchen Sie in der Datei *modelersrv.sh* die folgende Zeile:

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

Fügen Sie den Schalter **-request_ssl_password** wie folgt hinzu:

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

Installieren von Verschlüsselung mit unbegrenzter Stärke

Bei der als Teil des Produkts ausgelieferten Java Runtime Environment ist Verschlüsselung mit US-Exportstärke aktiviert. Zur besseren Sicherheit Ihrer Daten wird ein Upgrade auf eine Verschlüsselung mit unbegrenzter Stärke empfohlen.

IBM J9

1. Laden Sie die Java Cryptography Extension (JCE) Unlimited Strength Jurisdiction Policy Files für Ihre Version des SDK von der Website IBM.com herunter.
2. Extrahieren Sie die in der komprimierten Datei gepackten Unlimited Jurisdiction Policy Files. Die komprimierte Datei enthält eine Datei namens *US_export_policy.jar* und eine Datei namens *local_policy.jar*. Wechseln Sie in Ihrer Installation von WebSphere Application Server zum Verzeichnis *\$JAVA_HOME/jre/lib/security* und erstellen Sie eine Sicherungskopie der Dateien *US_export_policy.jar* und *local_policy.jar*.
3. Ersetzen Sie die vorhandenen Dateien *US_export_policy.jar* und *local_policy.jar* durch die beiden heruntergeladenen Dateien.
4. Aktivieren Sie die Sicherheit in der Administrationskonsole von WebSphere Application Server. Vergewissern Sie sich vorab, dass alle Knotenagenten in der Zelle aktiv sind. Weitere Informationen finden Sie in der WebSphere-Dokumentation. Beachten Sie, dass Sie eine verfügbare Realmdefinition aus der Liste unter **Sicherheit > Sichere Verwaltung, Anwendungen und Infrastruktur** auswählen müssen, und klicken Sie dann auf **Als aktuell festlegen**, sodass die Sicherheit bei einem Serverneustart aktiviert wird.
5. Melden Sie sich bei der Administrationskonsole ab.

6. Stoppen Sie den Server.
7. Starten Sie den Server neu.

Sun Java

1. Laden Sie die Java Cryptography Extension (JCE) Unlimited Strength Jurisdiction Policy Files für Ihre Version des SDK von der Sun Java-Website herunter.
2. Extrahieren Sie die heruntergeladene Datei.
3. Kopieren Sie die beiden JAR-Dateien *local_policy.jar* und *US_export_policy.jar* in das Verzeichnis *<Installationsordner>/jre/lib/security*, wobei *<Installationsordner>* der Ordner ist, in dem Sie das Produkt installiert haben.

Anweisung an Benutzer, SSL zu aktivieren

Wenn Endbenutzer über ein Clientprodukt eine Verbindung zum Server herstellen, müssen Sie SSL im Dialogfeld für die Verbindung zum Server aktivieren. Fordern Sie Ihre Benutzer unbedingt auf, das korrekte Kontrollkästchen zu markieren. In IBM SPSS Modeler beispielsweise sollten die Benutzer das Kontrollkästchen **Verbindung verschlüsseln** im Dialogfeld "Repository: Server" aktivieren, das über die Option **Repository** im Menü **Extras** aufgerufen werden kann.

Kapitel 5. Leistungsüberblick

Die tatsächliche Leistung beim Analysieren von Daten wird durch eine Reihe an Faktoren beeinflusst, von der Server- und Datenbankkonfiguration bis hin zum Sortieren von einzelnen Knoten in einem Stream. Im Allgemeinen erzielen Sie die beste Leistung, indem Sie die folgenden Punkte beachten:

- Speichern Sie Ihre Daten in einem DBMS und verwenden Sie die SQL-Generierung, wann immer dies möglich ist.
- Verwenden Sie Hardware, die die Empfehlungen unter Kapitel 2, „Architektur- und Hardwareempfehlungen“, auf Seite 7 erfüllt oder übertrifft.
- Stellen Sie sicher, dass die Leistungs- und Optimierungseinstellungen für Client und Server ordnungsgemäß konfiguriert sind. Beachten Sie: Wenn SPSS Modeler mit einer SPSS Modeler Server-Installation verbunden ist, setzen die Leistungs- und Optimierungseinstellungen des Servers die entsprechenden Einstellungen auf dem Client außer Kraft.
- Konzipieren Sie Streams für maximale Leistung.

In den folgenden Abschnitten finden Sie Informationen zu jedem einzelnen dieser Leistungsfaktoren.

Leistungs- und Optimierungseinstellungen für Ihren Server

Bestimmte IBM SPSS Modeler Server-Einstellungen können konfiguriert werden, um die Leistung zu optimieren. Sie können diese Einstellungen mithilfe der Anwendung IBM SPSS Modeler Administration Console anpassen, die in IBM SPSS Modeler Server enthalten ist. Weitere Informationen finden Sie im Thema „Verwenden von IBM SPSS Modeler Administration Console“ auf Seite 23.

Diese Einstellungen sind unter der Überschrift **Leistung und Optimierung** im Konfigurationsfenster von IBM SPSS Modeler Administration Console zusammengefasst. Die Einstellungen sind bei den meisten Installationen für die optimale Leistung vorkonfiguriert. Je nach Hardware, Dataset und Streaminhalt kann es erforderlich sein, diese Einstellungen anzupassen. Weitere Informationen finden Sie im Thema „Leistung und Optimierung“ auf Seite 26.

Leistungs- und Optimierungseinstellungen für Ihren Client

Die Leistungs- und Optimierungseinstellungen für den Client können über die Registerkarte "Optionen" des Dialogfelds "Streameigenschaften" aufgerufen werden. Wählen Sie zur Anzeige dieser Optionen Folgendes aus dem Client-Menü aus:

Tools > Streameigenschaften > Optionen > Optimierung

Sie können die Optimierungseinstellungen verwenden, um die Streamleistung zu optimieren. Beachten Sie, dass die Leistungs- und Optimierungseinstellungen auf IBM SPSS Modeler Server (sofern verwendet) alle entsprechenden Einstellungen auf dem Client außer Kraft setzen.

Hinweis: Datenbankmodellierung und SQL-Optimierung erfordern, dass auf dem IBM SPSS Modeler-Computer IBM SPSS Modeler Server-Konnektivität aktiviert ist. Wenn diese Einstellung aktiviert ist, können Sie auf Datenbankalgorithmen zugreifen, SQL direkt aus IBM SPSS Modeler per Pushback übertragen und auf IBM SPSS Modeler Server zugreifen. Wählen Sie zur Überprüfung des aktuellen Lizenzstatus die folgenden Optionen aus dem IBM SPSS Modeler-Menü aus.

Hilfe > Info > Zusätzliche Details

Wenn Konnektivität aktiviert ist, wird auf der Registerkarte "Lizenzstatus" die Option **Serveraktivierung** angezeigt.

Weitere Informationen finden Sie im Thema „Verbinden mit IBM SPSS Modeler Server“ auf Seite 15.

Hinweis: Die Unterstützung für SQL-Pushback ist von der verwendeten Datenbanktyp abhängig. Neueste Informationen zu Datenbanken und ODBC-Treibern, die für die Verwendung mit IBM SPSS Modeler 16 unterstützt werden und getestet wurden, finden Sie auf der unternehmensweiten Support-Site unter <http://www.ibm.com/support>.

Streamneuerstellung aktivieren. Wählen Sie diese Option aus, um die Streamneuerstellung in IBM SPSS Modeler zu aktivieren. Es stehen vier Neuerstellungsarten zur Verfügung. Sie können eine oder mehrere auswählen. Die Streamneuerstellung ordnet die Knoten in einem Stream für eine effizientere Operation im Hintergrund neu an, ohne die Streamsemantik zu ändern.

- **SQL-Generierung optimieren.** Mit dieser Option können die Knoten im Stream so neu angeordnet werden, dass ein Pushback weiterer Operationen mithilfe der SQL-Generierung für die Ausführung in der Datenbank möglich ist. Wenn die Optimierungsfunktion einen Knoten findet, der nicht in SQL wiedergegeben werden kann, sucht sie nachgeordnete Knoten, die in SQL wiedergegeben und ohne Beeinträchtigung der Streamsemantik vor den problematischen Knoten verschoben werden können. Nicht nur, dass die Datenbank die Operationen effizienter als IBM SPSS Modeler durchführen kann, solche Pushbacks dienen auch dazu, die Größe des Datensets, das an IBM SPSS Modeler zur Verarbeitung zurückgegeben wird, zu reduzieren. So wird wiederum der Netzverkehr reduziert und die Geschwindigkeit der Streamoperationen erhöht. Das Kontrollkästchen **SQL generieren** muss aktiviert sein, damit die SQL-Optimierung funktionieren kann.
- **CLEM-Ausdruck optimieren.** Mit dieser Option kann das Optimierungsprogramm nach CLEM-Ausdrücken suchen, die vor der Ausführung des Streams vorverarbeitet werden können, damit die Verarbeitungsgeschwindigkeit erhöht wird. Ein einfaches Beispiel dafür ist, dass das Optimierungsprogramm bei einem Ausdruck wie $\log(\text{Gehalt})$ den tatsächlichen Gehaltswert berechnen und diesen dann für die Verarbeitung übergeben würde. Dies kann die Leistung von SQL-Pushback und IBM SPSS Modeler Server erhöhen.
- **Syntaxausführung optimieren.** Diese Methode der Streamneuerstellung erhöht die Effizienz von Operationen, die mehr als einen Knoten mit IBM SPSS Statistics-Syntax umfassen. Die Optimierung wird durch die Zusammenfassung der Syntaxbefehle in einer einzigen Operation erreicht, anstatt jeden einzelnen Syntaxbefehl als separate Operation auszuführen.
- **Andere Ausführung optimieren.** Diese Methode der Streamneuerstellung erhöht die Effizienz von Operationen, die nicht an die Datenbank delegiert werden können. Eine Optimierung wird dadurch erreicht, dass die Datenmenge im Stream so früh wie möglich reduziert wird. Während die Datenintegrität gewahrt bleibt, wird der Stream neu erstellt, um die Operationen näher zur Datenquelle zu bringen. So werden die Daten abwärts für kostenintensive Operationen, z. B. Joins, reduziert.

Parallele Verarbeitung aktivieren. Bei Ausführung auf einem Computer mit mehreren Prozessoren ermöglicht diese Option dem System, die Last auf diese Prozessoren zu verteilen, was zu einer schnelleren Leistung führen kann. Bei Verwendung mehrerer Knoten oder der folgenden Einzelknoten können Sie von der parallelen Verarbeitung profitieren: C5.0, Zusammenführen (nach Schlüssel), Sortieren, Klasse (Rangmethode und N-Perzentil-Methode) sowie Aggregieren (unter Verwendung mindestens eines Schlüsselfelds).

SQL generieren. Wählen Sie diese Option aus, um die SQL-Generierung zu aktivieren. Dabei wird zugelassen, dass Streamoperationen per Pushback an die Datenbank zurückübertragen werden, indem SQL-Code zur Generierung der Ausführungsprozesse verwendet wird, was zu einer Leistungssteigerung führen kann. Zur weiteren Steigerung der Leistung kann außerdem **SQL-Generierung optimieren** ausgewählt werden. Dadurch wird die Anzahl der Operationen, die per Pushback an die Datenbank zurückübertragen werden, maximiert. Wenn Operationen für einen Knoten per Pushback an die Datenbank zurückübertragen wurden, wird der Knoten bei der Ausführung des Streams violett gekennzeichnet.

- **Datenbankcaching.** Bei Streams, die in der Datenbank auszuführendes SQL generieren, muss nicht das Dateisystem als Cache verwendet werden, sondern die Daten können mitten im Stream in einer temporären Tabelle in der Datenbank zwischengespeichert werden. Bei einer Kombination mit SQL-Optimierung kann dies zu einer erheblichen Leistungssteigerung führen. So kann beispielsweise die Aus-

gabe aus einem Stream, der mehrere Tabellen zusammenfasst, um eine Data-Mining-Ansicht zu erstellen, im Cache gespeichert und bei Bedarf wieder verwendet werden. Klicken Sie bei aktiviertem Datenbankcaching einfach mit der rechten Maustaste auf einen Knoten (keinen Endknoten), um die Daten an dieser Stelle im Cache zu speichern, und der Cache wird bei der nächsten Ausführung des Streams automatisch unmittelbar in der Datenbank erstellt. Dadurch kann SQL für nachgeordnete Knoten generiert werden, was zu weiteren Leistungsverbesserungen führt. (Alternativ kann diese Option bei Bedarf inaktiviert werden, beispielsweise wenn es aufgrund von Richtlinien oder Berechtigungen nicht möglich ist, Daten in die Datenbank zu schreiben. Wenn Datenbankcaching oder SQL-Optimierung nicht aktiviert ist, wird der Cache stattdessen in das Dateisystem geschrieben.

- **Unproblematische Umwandlung verwenden.** Mit dieser Option können Daten von Zeichenfolgen in Zahlen oder von Zahlen in Zeichenfolgen konvertiert werden, sofern sie in einem geeigneten Format gespeichert sind. Wenn die Daten beispielsweise als Zeichenfolge in der Datenbank gespeichert sind, tatsächlich jedoch eine sinnvolle Zahl enthalten, können die Daten beim Eintreten des Pushbacks zur Verwendung konvertiert werden.

Hinweis: Aufgrund kleinerer Unterschiede in der SQL-Implementierung weichen die Ergebnisse der in einer Datenbank ausgeführten Streams möglicherweise ein wenig von den bei der Ausführung in IBM SPSS Modeler gelieferten Ergebnissen ab. Aus ähnlichen Gründen können diese Unterschiede auch je nach Datenbankanbieter variieren.

Datenbankverwendung und -optimierung

Datenbankserver. Wenn möglich, erstellen Sie eine dedizierte Datenbankinstanz für das Data-Mining, so dass der Produktionsserver nicht von IBM SPSS Modeler-Anfragen beeinflusst wird. Die von IBM SPSS Modeler generierten SQL-Anweisungen können einen großen Teil der Ressourcen in Anspruch nehmen. Auf dem IBM SPSS Modeler Server-Computer können mehrere Aufgaben SQL in derselben Datenbank ausführen.

Datenbankinternes Mining. Viele Datenbankanbieter bieten für ihre Produkte Data-Mining-Erweiterungen an. Mit diesen Erweiterungen können Sie Data-Mining-Aufgaben (z. B. Modellerstellung oder Scoring) auf dem Datenbankserver oder einem separaten dedizierten Server ausführen. Die datenbankinternen Mining-Funktionen von IBM SPSS Modeler ergänzen und erweitern die SQL-Generierungsfähigkeiten des Produkts und bieten auf diese Weise eine Möglichkeit, die herstellerspezifischen Datenbankerweiterungen zu betreiben. In manchen Fällen kann diese Herangehensweise den manchmal kostenintensiven Aufwand für die Datenübertragung zwischen IBM SPSS Modeler und der Datenbank reduzieren. Auch das Datenbankcaching kann dazu beitragen, diese Vorteile weiter auszubauen. Weitere Informationen finden Sie in der Datei *DatabaseMiningGuide.pdf* im Ordner *\Documentation* auf der IBM SPSS Modeler-Installations-CD.

SQL-Optimierung

Die beste Leistung erzielen Sie, wenn Sie die generierte Menge an SQL voll ausschöpfen und somit die Leistungsfähigkeit und Skalierbarkeit der Datenbank nutzen. Nur die Teile des Streams, die nicht in SQL kompiliert werden können, sollten in IBM SPSS Modeler Server ausgeführt werden. Weitere Informationen finden Sie in Kapitel 6, „SQL-Optimierung“, auf Seite 39.

Hochladen von dateibasierten Daten

Daten, die nicht in einer Datenbank gespeichert sind, können nicht von der SQL-Optimierung profitieren. Wenn sich die zu analysierenden Daten noch nicht in einer Datenbank befinden, können Sie sie mit einem Datenbankausgabeknoten hochladen. Sie können diesen Knoten außerdem verwenden, um Datenzwischenmengen aus der Datenvorbereitung und die Ergebnisse der Bereitstellung zu speichern.

IBM SPSS Modeler kann mit den externen Ladeprogrammen für viele gängige Datenbanksysteme zusammenarbeiten. In der Software sind mehrere Scripts enthalten. Diese stehen (mit Dokumentation) im Unterverzeichnis */scripts* unter dem IBM SPSS Modeler-Installationsordner zur Verfügung.

In der folgenden Tabelle wird der potenzielle Leistungsvorteil des Massenladens gezeigt. In der Abbildung wird dargestellt, wie viel Zeit zum Exportieren von 250.000 Datensätzen und 21 Feldern in eine Oracle-Datenbank erforderlich ist. Das externe Ladeprogramm ist das Oracle-Dienstprogramm sqlldr.

Tabelle 2. Leistungsvorteil des Massenladens

Exportoption	Zeit (in Sekunden)
Standard (ODBC)	409
Massenladen über ODBC	52
Massenladen über externes Ladeprogramm	33

Kapitel 6. SQL-Optimierung

Eine der wichtigsten Funktionen von IBM SPSS Modeler besteht in der Möglichkeit, mehrere Datenvorbereitungs- und Mining-Operationen direkt in der Datenbank durchzuführen. Durch die Generierung von SQL-Code, der für die Ausführung zurück in die Datenbank übertragen wird (Pushback), können viele Operationen, wie die Stichprobenziehung, Sortierung, Ableitung neuer Felder und bestimmte Formen der Diagrammerstellung, anstatt auf dem IBM SPSS Modeler- oder IBM SPSS Modeler Server-Computer in der Datenbank durchgeführt werden. Wenn Sie mit großen Datasets arbeiten, können diese **Pushbacks** die Leistung auf unterschiedliche Weise erheblich steigern.

- Durch Reduzierung der Größe des Ergebnisses, das aus dem DBMS nach IBM SPSS Modeler übertragen werden soll. Wenn große Ergebnisses mithilfe eines ODBC-Treibers gelesen werden, kann es zu Netz-E/A- oder Treiberfehlern kommen. Aus diesem Grund sind Zeilen- und Spaltenauswahl und -aggregation (Auswahl-, Stichproben-, Aggregatknoten), mit denen üblicherweise die Größe des zu übertragenden Datasets verringert wird, die Operationen, die am meisten von der SQL-Optimierung profitieren. Daten können außerdem an kritischen Punkten im Stream (z. B. nach einem Merge- oder Auswahlknoten) in einer temporären Tabelle in der Datenbank zwischengespeichert werden, um die Leistung weiter zu steigern.
- Durch Nutzung der Leistung und Skalierbarkeit der Datenbank. Die Effizienz wird erhöht, da ein DBMS häufig die Vorteile paralleler Verarbeitung, leistungsstärkerer Hardware, einer ausgereifteren Verwaltung des Speicherplatzes und des Vorhandenseins von Indizes nutzen kann.

Angesichts dieser Vorteile wurde IBM SPSS Modeler so entworfen, dass es die Menge an SQL, das von den einzelnen Streams generiert wird, maximiert, sodass nur Operationen, die sich nicht in SQL kompilieren lassen, von IBM SPSS Modeler Server ausgeführt werden. Aufgrund der Beschränkungen der Ausdrucksmöglichkeiten in der Standard-SQL (SQL-92) werden bestimmte Operationen nicht unterstützt. Weitere Informationen finden Sie im Thema „Tipps für die Maximierung der SQL-Generierung“ auf Seite 43.

Hinweis: Aufgrund kleinerer Unterschiede in der SQL-Implementierung weichen die Ergebnisse der in einer Datenbank ausgeführten Streams möglicherweise ein wenig von den bei der Ausführung in IBM SPSS Modeler erzielten Ergebnissen ab. Diese Unterschiede können aus ähnlichen Gründen auch je nach Datenbankanbieter variieren. Je nach der Datenbankkonfiguration für die Unterscheidung zwischen Groß- und Kleinschreibung beim Vergleich und der Sortierung von Zeichenfolgen können IBM SPSS Modeler-Streams, die unter Verwendung von SQL-Pushback ausgeführt werden, zu anderen Ergebnissen führen als Streams, die ohne SQL-Pushback ausgeführt werden. Wenden Sie sich an den Datenbankadministrator, wenn Sie Unterstützung bei der Konfiguration Ihrer Datenbank benötigen. Für größtmögliche Kompatibilität mit IBM SPSS Modeler sollte beim Vergleich von Datenbankzeichenfolgen zwischen Groß- und Kleinschreibung unterschieden werden.

Hinweis: Datenbankmodellierung und SQL-Optimierung erfordern, dass auf dem IBM SPSS Modeler-Computer IBM SPSS Modeler Server-Konnektivität aktiviert ist. Wenn diese Einstellung aktiviert ist, können Sie auf Datenbankalgorithmen zugreifen, SQL direkt aus IBM SPSS Modeler per Pushback übertragen und auf IBM SPSS Modeler Server zugreifen. Wählen Sie zur Überprüfung des aktuellen Lizenzstatus die folgenden Optionen aus dem IBM SPSS Modeler-Menü aus.

Hilfe > Info > Zusätzliche Details

Wenn Konnektivität aktiviert ist, wird auf der Registerkarte "Lizenzstatus" die Option **Serveraktivierung** angezeigt.

Weitere Informationen finden Sie im Thema „Verbinden mit IBM SPSS Modeler Server“ auf Seite 15.

Hinweis: Wenn Streams in einer Netezza-Datenbank ausgeführt werden, werden die Angaben zu Datum und Uhrzeit aus der betreffenden Datenbank übernommen. Diese Werte können von den lokalen Werten für Datum und Uhrzeit bzw. den Werten in IBM SPSS Modeler Server abweichen, wenn sich die Datenbank beispielsweise auf einem Computer in einem anderen Land bzw. einer anderen Zeitzone befindet.

Datenbankanforderungen

Neueste Informationen zu Datenbanken und ODBC-Treibern, die für die Verwendung mit IBM SPSS Modeler unterstützt werden und getestet wurden, finden Sie in den Produktkompatibilitätsdiagrammen auf der unternehmensweiten Support-Site unter <http://www.ibm.com/support>.

Beachten Sie, dass durch Verwendung von Datenbankmodellierung eventuell weitere Leistungsverbesserungen möglich sind.

Funktionsweise der SQL-Generierung

Die ersten Fragmente eines Streams, der von den Quellenknoten der Datenbank ausgeht, sind die Hauptziele für die SQL-Generierung. Wenn ein Knoten gefunden wird, der sich nicht in SQL kompilieren lässt, werden die Daten aus der Datenbank extrahiert und die weitere Verarbeitung wird von IBM SPSS Modeler Server durchgeführt.

Während der Streamvorbereitung und vor der Ausführung findet die SQL-Generierung wie folgt statt.

- Der Server ordnet Streams neu, um nachgeordnete Knoten in die "SQL-Zone" zu verschieben, wenn dies erwiesenermaßen problemlos möglich ist. (Diese Funktion kann auf dem Server inaktiviert werden.)
- SQL-Ausdrücke werden schrittweise von den Quellenknoten zu den Endknoten hin erstellt. Diese Phase endet, wenn ein Knoten gefunden wird, der nicht in SQL konvertiert werden kann, oder der Endknoten (z. B. Tabellenknoten oder Diagrammknoten) in SQL konvertiert wird. Am Ende dieser Phase wird jeder Knoten mit einer SQL-Anweisung beschriftet, falls der Knoten und seine Vorgänger eine SQL-Entsprechung aufweisen.
- Das SQL wird ausgehend von den Knoten mit den kompliziertesten SQL-Äquivalenten zurück zu den Quellenknoten auf Validität überprüft. Das SQL, das erfolgreich validiert wurde, wird für die Ausführung ausgewählt.
- Knoten, für die alle Operationen SQL generiert haben, werden im Streamerstellungsbereich mit violetter Markierung angezeigt. Auf der Grundlage der Ergebnisse kann es angebracht sein, den Stream an den entsprechenden Stellen weiter zu reorganisieren, um die Datenbankausführung bestmöglich zu nutzen. Weitere Informationen finden Sie im Thema „Tipps für die Maximierung der SQL-Generierung“ auf Seite 43.

Auftreten von Verbesserungen

SQL-Optimierung verbessert die Leistung in einer Reihe von Datenoperationen:

- **Joins (Zusammenführen (Merge) nach Schlüssel).** Joinoperationen können die Optimierung innerhalb von Datenbanken steigern.
- **Aggregation.** Die Aggregat-, Verteilungs- und Netzdiagrammknoten verwenden jeweils Aggregation zum Erstellen der Ergebnisse. Zusammengefasste Daten benötigen erheblich weniger Bandbreite als die ursprünglichen Daten.
- **Auswahl.** Die Auswahl von Datensätzen auf der Grundlage bestimmter Kriterien verringert die Menge der Datensätze.
- **Sortieren.** Das Sortieren von Datensätzen ist eine ressourcenintensive Aktivität, die sich effizienter in einer Datenbank ausführen lässt.
- **Feldableitung.** Neue Felder lassen sich effizienter in einer Datenbank generieren.

- **Feldprojektion.** IBM SPSS Modeler Server extrahiert nur Felder aus der Datenbank, die für die nachfolgende Verarbeitung erforderlich sind. Dadurch werden die Anforderungen an Bandbreite und Arbeitsspeicher minimiert. Dasselbe gilt auch für überflüssige Felder in Flatfiles: Der Server muss die überflüssigen Felder zwar lesen, ordnet ihnen jedoch keinen Speicher zu.
- **Scoring.** SQL lässt sich aus Entscheidungsbäumen, Regelsets, linearer Regression und faktorgenerierten Modellen generieren.

Beispiel für SQL-Generierung

Mit dem folgenden Stream werden drei Datenbanktabellen mithilfe von Schlüsseloperationen zusammengefügt und dann sowohl eine Aggregation als auch eine Sortierung durchgeführt.

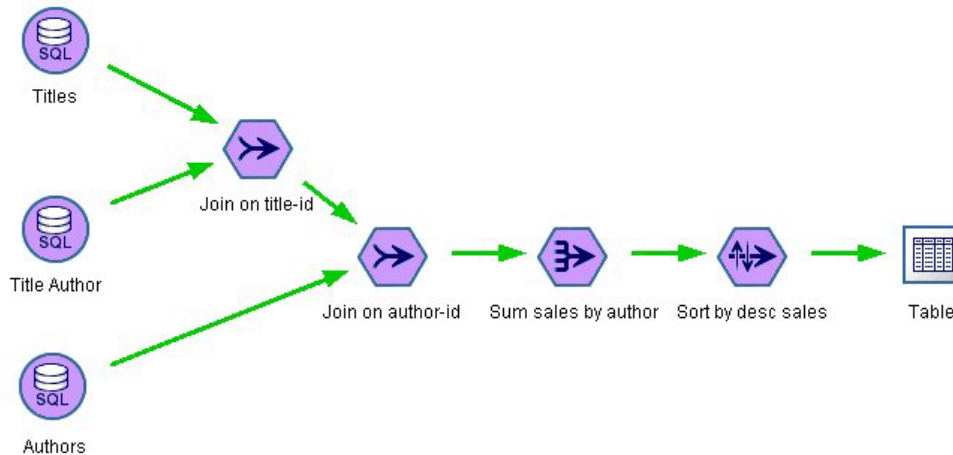


Abbildung 3. Optimierter Stream mit violetten Knoten, die SQL-Pushbacks (in der Datenbank durchgeführte Operationen) anzeigen

Generiertes SQL

Das für diesen Stream generierte SQL lautet wie folgt:

```
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
```

Ausführen des Streams

Wenn der Stream mit einem Datenbankexportknoten beendet wird, kann der gesamte Stream in der Datenbank ausgeführt werden.

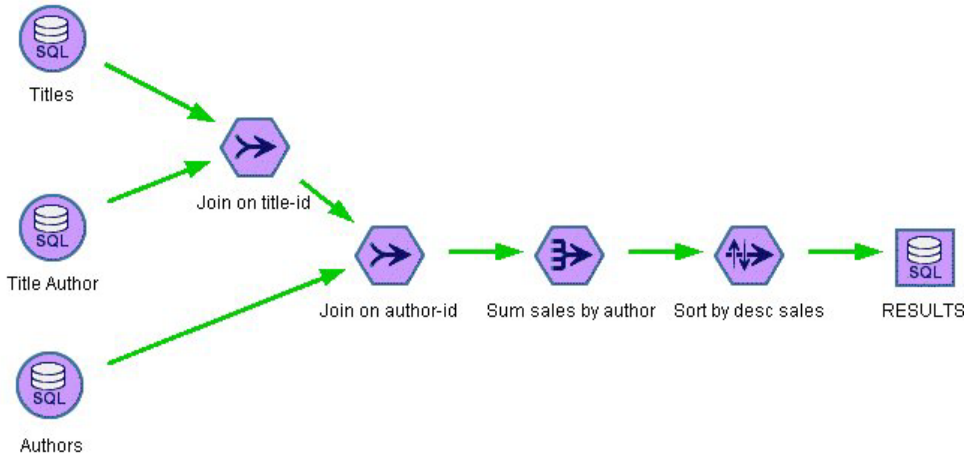


Abbildung 4. Gesamter, in der Datenbank ausgeführter Stream

Konfigurieren der SQL-Optimierung

1. Installieren Sie einen ODBC-Treiber und konfigurieren Sie eine Datenquelle für die zu verwendende Datenbank. Weitere Informationen finden Sie im Thema „Datenzugriff“ auf Seite 11.
2. Erstellen Sie einen Stream, der einen Quellenknoten zum Extrahieren von Daten aus dieser Datenbank verwendet.
3. Vergewissern Sie sich gegebenenfalls, dass die SQL-Generierung auf Client und Server aktiviert ist. Standardmäßig ist sie für beide aktiviert.

So aktivieren Sie SQL-Optimierung auf dem Client

1. Wählen Sie im Menü "Extras" die Option **Streameigenschaften > Optionen** aus.
2. Klicken Sie dann auf die Registerkarte "Optimierung". Wählen Sie **SQL generieren**, um die SQL-Optimierung zu aktivieren. Optional können Sie andere Einstellungen auswählen, um die Leistung zu verbessern. Weitere Informationen finden Sie im Thema „Leistungs- und Optimierungseinstellungen für Ihren Client“ auf Seite 35.

So aktivieren Sie SQL-Optimierung auf dem Server

Da die Servereinstellungen alle auf dem Client angegebenen Spezifikationen überschreiben, müssen die Serverkonfigurationseinstellungen **Streamneuerstellung** und **Automatische SQL-Generierung** jeweils aktiviert sein. Weitere Informationen zum Ändern der IBM SPSS Modeler Server-Einstellungen finden Sie im Abschnitt „Leistung und Optimierung“ auf Seite 26.

So aktivieren Sie die Optimierung beim Scoring von Modellen

Zum Zwecke des Scorings muss die SQL-Generierung für jeden Modellierungsknoten separat aktiviert werden, unabhängig von den Einstellungen auf der Server- bzw. Client-Ebene. Das liegt daran, dass manche Modelle extrem komplexe SQL-Ausdrücke generieren, die in der Datenbank unter Umständen nicht effektiv evaluiert werden. In der Datenbank können Fehler auftreten, wenn das generierte SQL ausgeführt werden soll, weil das SQL zu groß oder zu komplex ist.

Unter Umständen erfordert es einige Anläufe, um festzustellen, ob die SQL-Generierung die Leistung eines bestimmten Modells verbessert. Dies geschieht auf der Registerkarte "Einstellungen", nachdem ein generiertes Modell zu einem Stream hinzugefügt wurde.

Vorschau für generiertes SQL

Sie können das generierte SQL im Nachrichtenprotokoll in der Vorschau anzeigen, bevor sie in der Datenbank ausgeführt wird. Dies ist u. U. für die Fehlersuche nützlich und ermöglicht außerdem den Export des generierten SQL, um diese zu einem späteren Zeitpunkt in der Datenbank zu bearbeiten bzw. auszuführen. Außerdem wird angezeigt, welche Knoten per Pushback an die Datenbank zurückübertragen werden. Dadurch können Sie leichter ermitteln, ob sich durch eine Neuordnung des Streams eine Leistungssteigerung erzielen lässt.

1. Vergewissern Sie sich, dass die Optionen **SQL während der Streamausführung im Nachrichtenprotokoll anzeigen** und **Einzelheiten zur SQL-Generierung während der Streamvorbereitung im Nachrichtenprotokoll anzeigen** im Dialogfeld "Benutzeroptionen" ausgewählt sind. Weitere Informationen finden Sie im Thema „Leistungs- und Optimierungseinstellungen für Ihren Client“ auf Seite 35.
2. Wählen Sie im Streamerstellungsbereich den Knoten oder Stream aus, den Sie in der Vorschau anzeigen möchten.
3. Klicken Sie in der Symbolleiste auf die Schaltfläche für die **SQL-Vorschau**.
Alle Knoten, für die SQL generiert wird (und die bei der Ausführung des Streams per Pushback in die Datenbank zurückübertragen werden) werden im Streamerstellungsbereich violett gekennzeichnet.
4. Um das generierte SQL in der Vorschau anzuzeigen, wählen Sie die folgende Optionsfolge aus den Menüs aus:
Extras > Streameigenschaften > Nachrichten...

Anzeigen von SQL für Modellnuggets

Bei einigen Modellen kann SQL für das Modellnugget generiert werden, wobei die Modellscoring-Phase per Pushback an die Datenbank zurückgegeben wird. Diese Funktion dient nicht in erster Linie zur Leistungsverbesserung, sondern dazu, dass das gesamte SQL von Streams, die diese Nuggets enthalten, per Pushback zurückgeführt werden kann. Weitere Informationen finden Sie im Thema „Knoten zur Unterstützung der SQL-Generierung“ auf Seite 44.

Führen Sie folgende Schritte aus, um das SQL für ein Modellnugget anzuzeigen, das SQL-Generierung unterstützt:

1. Wählen Sie die Registerkarte "Einstellungen" auf dem Modellnugget.
2. Wählen Sie nach Bedarf eine der Optionen **Mit (Ohne) Unterstützung für fehlende Werte generieren** oder **SQL für dieses Modell generieren**.
3. Wählen Sie im Menü des Modellnuggets:
Datei > SQL exportieren
4. Speichern Sie die Datei.
5. Öffnen Sie die Datei, um das SQL anzuzeigen.

Tipps für die Maximierung der SQL-Generierung

Um die bestmögliche Leistungssteigerung aus der SQL-Optimierung zu erzielen, sollten Sie auf folgende Punkte achten:

Streamreihenfolge. Die SQL-Generierung wird möglicherweise angehalten, wenn die Funktion des Knotens keine semantische Entsprechung in SQL aufweist, da die Data-Mining-Funktion von IBM SPSS Modeller umfassender ist als die herkömmlichen Datenverarbeitungsoperationen, die von Standard-SQL unterstützt werden. In diesem Fall wird die SQL-Generierung auch für alle nachgeordneten Knoten unterdrückt. Daher können Sie die Leistung eventuell signifikant steigern, wenn Sie die Knoten neu anordnen, um die Operationen, die SQL stoppen, so weit unten im Stream platzieren wie möglich. Das SQL-Optimierungsprogramm kann Neuordnungen bis zu einem bestimmten Grad automatisch durchführen (sie müssen jedoch sicherstellen, dass die Streamneuerstellung aktiviert ist), es sind jedoch eventuell

weitere Verbesserungen möglich. Ein guter Kandidat dafür ist der Auswahlknoten, der häufig nach weiter vorne im Stream verschoben werden kann. Weitere Informationen finden Sie im Thema „Knoten zur Unterstützung der SQL-Generierung“.

CLEM-Ausdrücke. Wenn ein Stream nicht neu geordnet werden kann, können Sie eventuell Knotenoptionen oder CLEM-Ausdrücke ändern oder anderweitig die Art und Weise, wie die Operation ausgeführt wird, umgestalten, sodass die SQL-Generierung nicht mehr behindert wird. Ableitungsknoten, Auswahlknoten und ähnliche Knoten können allgemein in SQL gerendert werden, vorausgesetzt, dass es für alle CLEM-Ausdrucks-Operatoren SQL-Entsprechungen gibt. Die meisten Operatoren können gerendert werden. Es gibt jedoch eine Reihe von Operationen, die die SQL-Generierung blockieren (insbesondere die Sequenzfunktionen [“@-Funktionen“]). Manchmal wird die Generierung angehalten, da die generierte Abfrage so komplex wurde, sodass die Datenbank nicht mehr damit umgehen kann. Weitere Informationen finden Sie im Thema „CLEM-Ausdrücke und Operatoren, die die SQL-Generierung unterstützen“ auf Seite 48.

Mehrere Quellenknoten. Wenn ein Stream mehrere Datenbankquellenknoten aufweist, wird die SQL-Generierung auf jeden Eingabezweig unabhängig angewendet. Wenn die Generierung in einem Zweig angehalten wird, kann sie in einem anderen fortgesetzt werden. Wenn zwei Zweige mit einem Merge-Vorgang zusammengeführt werden (und beide Zweige bis zum Merge-Punkt in SQL ausgedrückt werden können), lässt sich die Zusammenführung oft durch einen Datenbank-Join ersetzen und die Generierung kann weiter unten im Stream fortgesetzt werden.

Datenbankalgorithmus. Die Modellschätzung wird fast immer auf IBM SPSS Modeler Server und nicht in der Datenbank durchgeführt, es sei denn, es werden datenbankeigene Algorithmen von Microsoft, IBM bzw. Oracle verwendet.

Scoring-Modelle. Für einige Modelle wird datenbankinternes Scoring durch Rendern des generierten Modells in SQL unterstützt. Einige Modelle generieren jedoch extrem komplexe SQL-Ausdrücke, die nicht immer effektiv in der Datenbank evaluiert werden. Aus diesem Grund muss die SQL-Generierung separat für die einzelnen Modellknoten aktiviert werden. Wenn sich herausstellt, dass der Modellknoten die SQL-Generierung behindert, wechseln Sie zur Registerkarte "Einstellungen" im Dialogfeld des Knotens und wählen Sie **SQL für dieses Modell generieren** aus (bei einigen Modellen stehen Ihnen möglicherweise weitere Optionen zur Steuerung der SQL-Generierung zur Verfügung. Führen Sie Tests aus, um sicherzustellen, dass die Option für Ihre Anwendung nützlich ist. Weitere Informationen finden Sie im Thema „Knoten zur Unterstützung der SQL-Generierung“.

Wenn Sie Knoten testen, um festzustellen, ob SQL-Generierung für Modelle effektiv arbeitet, sollten Sie zuvor alle Streams von IBM SPSS Modeler speichern. Einige Datenbanksysteme können hängen bleiben, wenn sie versuchen, das (möglicherweise komplexe) generierte SQL zu verarbeiten, und es kann erforderlich sein, IBM SPSS Modeler über den Windows Task-Manager zu schließen.

Datenbankcaching. Wenn Sie einen Knotencache zum Speichern von Daten an kritischen Punkten im Stream verwenden (beispielsweise nach einem Merge- oder Aggregatknoten), müssen Sie sicherstellen, dass neben der SQL-Optimierung auch Datenbankcaching aktiviert ist. Dadurch können die Daten in den meisten Fällen eine temporäre Tabelle in der Datenbank (statt des Dateisystems) als Cache fungieren. Weitere Informationen finden Sie im Thema „Konfigurieren der SQL-Optimierung“ auf Seite 42.

Herstellerspezifisches SQL. Der größte Teil des generierten SQL entspricht den Standards (SQL-92), es werden jedoch, wo angemessen, auch einige nicht standardmäßige, herstellerspezifische Funktionen genutzt. Der Grad der SQL-Optimierung kann, je nach Datenbankquelle, schwanken.

Knoten zur Unterstützung der SQL-Generierung

In den folgenden Tabellen werden die Knoten für die Data-Mining-Operationen beschrieben, die eine SQL-Generierung unterstützen. Mit Ausnahme der Datenbankmodellknoten wird die SQL-Generierung von Knoten, die nicht in diesen Tabellen aufgeführt sind, nicht unterstützt.

Sie können das zu generierende SQL anzeigen, bevor Sie es ausführen. Weitere Informationen finden Sie im Thema „Vorschau für generiertes SQL“ auf Seite 43.

Tabelle 3. Quellen

Knoten zur Unterstützung der SQL-Generierung	Anmerkungen
Datenbank	Dieser Knoten wird verwendet, um Tabellen und Ansichten für eine weitere Analyse anzugeben. Dieser Knoten ermöglicht die Eingabe von SQL-Abfragen. Vermeiden Sie Ergebnissets mit doppelten Spaltennamen. Weitere Informationen finden Sie im Thema „Schreiben von SQL-Abfragen“ auf Seite 51.

Tabelle 4. Datensatzoperationen

Knoten zur Unterstützung der SQL-Generierung	Anmerkungen
Auswählen	Unterstützt die Generierung nur, wenn SQL-Generierung für den select-Ausdruck selbst unterstützt wird (siehe Ausdrücke unten). Wenn Felder mit Nullwerten vorliegen, führt die SQL-Generation für "discard" nicht zu denselben Ergebnissen wie in IBM SPSS Modeler.
Stichprobe	Die einfache Stichprobenziehung unterstützt die SQL-Generierung bis zu einem gewissen Grad je nach Datenbank. Siehe Tabelle 5.
Aggregieren	Die Unterstützung der SQL-Generierung für die Aggregation hängt vom Datenspeichertyp ab. Siehe Tabelle 6 auf Seite 46.
RFM-Aggregat	Unterstützt die Generierung, außer beim Speichern des Datums der zweit- oder drittaktuellsten Transaktion oder wenn nur aktuelle Transaktionen eingeschlossen werden. Die Aufnahme aktueller Transaktionen ist jedoch möglich, wenn für die Funktion <code>datetime_date(JAHR,MONAT,TAG)</code> ein Pushback-Vorgang durchgeführt wird.
Sortieren	
Zusammenführen	Keine SQL generiert für Zusammenführen (Merge) nach Schlüssel . Ein Zusammenführen nach Schlüssel mit Full oder partiellem Outer Join wird nur unterstützt, falls sie auch von der Datenbank/dem Treiber unterstützt wird. Nicht übereinstimmende Eingabefelder können mittels eines Filterknotens oder über die Registerkarte "Filter" eines Quellenknotens umbenannt werden. Unterstützt SQL-Generierung für die Zusammenführung nach Bedingung. Für alle Merge-Typen wird <code>SQL_SP_EXISTS</code> nicht unterstützt, wenn die Eingaben aus verschiedenen Datenbanken stammen.
Anhängen	Unterstützt die Generierung bei unsortierten Eingaben.
Duplikat	

Tabelle 5. Unterstützung der SQL-Generierung im Stichprobenknoten für einfache Stichprobenziehung.

Modalwert	Stichprobe	Max. Größe	Startwert	DB2 for z/OS	DB2 for OS/400	DB2 for Windows/UNIX	Netezza (siehe Hinweis)	Oracle	SQL Server	Teradata
Einschließen	Erster	entfällt		J	J	J	J	J	J	J
	1 in n	Aus		J	J	J	J	J		J
		max			J	J	J	J	J	

Tabelle 5. Unterstützung der SQL-Generierung im Stichprobenknoten für einfache Stichprobenziehung (Forts.).

Modalwert	Stichprobe	Max. Größe	Startwert	DB2 for z/OS	DB2 for OS/400	DB2 for Windows/UNIX	Netezza (siehe Hinweis)	Oracle	SQL Server	Teradata	
	Zufällig %	Aus	Aus			J	J	J		J	
			Ein			J		J			
		max	Aus			J	J	J		J	
			Ein			J		J			
Verwerfen	Erster	Aus					J	J			
		max					J	J			
	1 in n	Aus		J	J	J	J	J		J	
		max		J	J	J	J	J		J	
	Zufällig %	Aus	Aus	Aus			J	J	J		J
			Ein			J		J			
		max	Aus			J	J	J		J	
			Ein			J		J			

Hinweis: Für Netezza-Datenbanken bezieht sich die Einschränkung auf die Verwendung des Schlüsselworts in Unteranfragen, die zur SQL-Generierung verwendet werden. Als Ausweidlösung kann Caching im Stichprobenknoten aktiviert werden.

Tabelle 6. Unterstützung der SQL-Generierung im Aggregatknoten.

Speicher	Summe	Mittelwert	Min	Max	Std. Abw.	Median	Anzahl	Varianz	Perzentil
Ganzzahl	J	J	J	J	J		J	J	
Reelle Zahl	J	J	J	J	J		J	J	
Datum			J	J			J		
Zeit			J	J			J		
Zeitmarke			J	J			J		
Zeichenfolge			J	J			J		

Tabelle 7. Feldoperationen

Knoten zur Unterstützung der SQL-Generierung	Anmerkungen
Typ	Unterstützt die SQL-Generierung, wenn der Typknoten instanziiert wird und keine ABORT- oder WARN-Typprüfung angegeben ist.
Filter	
Ableiten	Unterstützt die SQL-Generierung, wenn das für den Ableitungsausdruck generierte SQL unterstützt wird (siehe Ausdrücke unten).
Ensemble	Unterstützt SQL-Generierung für stetige Ziele. Bei anderen Zielen wird die Generierung nur unterstützt, wenn die Ensemble-Methode "Höchste Konfidenz hat Vorrang" verwendet wird.
Füller	Unterstützt die SQL-Generierung, wenn das für den "derive"-Ausdruck generierte SQL unterstützt wird (siehe Ausdrücke unten).

Tabelle 7. Feldoperationen (Forts.)

Knoten zur Unterstützung der SQL-Generierung	Anmerkungen
Anonymisieren	Unterstützt die SQL-Generierung für stetige Ziele und die partielle SQL-Generierung für nominale und Flagziele.
Umcodieren	
Klassierung	Unterstützt SQL-Generierung, wenn die Klassiermethode "n-tile (gleiche Anzahl)" verwendet wird und die Option "Von Registerkarte "Klassenwerte" lesen, sofern verfügbar" ausgewählt ist.
RFM-Analyse	Unterstützt SQL-Generierung, wenn die Option "Von Registerkarte "Klassenwerte" lesen, sofern verfügbar" ausgewählt ist, jedoch unterstützen dies nachgeordnete Knoten nicht.
Partitionieren	Unterstützt die SQL-Generierung für die Zuweisung von Datensätzen zu Partitionen.
Dichotom	
Umstrukturieren	

Tabelle 8. Grafiken

Knoten zur Unterstützung der SQL-Generierung	Anmerkungen
Diagrammtafel	SQL-Generierung wird für die folgenden Diagrammtypen unterstützt: Fläche, 3-D-Fläche, Balken, 3-D-Balken, Balken für Häufigkeiten, Heat-Map, Kreis, 3-D-Kreis, Kreis für Häufigkeiten. Bei Histogrammen wird SQL-Generierung nur für kategoriale Daten unterstützt.
Verteilung	
Internet	
Evaluation	

Tabelle 9. Modellierung

Knoten zur Unterstützung der SQL-Generierung	Anmerkungen
Automatisches Klassifikationsmerkmal	Unterstützt SQL-Generierung für stetige Ziele. Bei anderen Zielen wird die Generierung nur unterstützt, wenn die Ensemble-Methode "Höchste Konfidenz hat Vorrang" verwendet wird.
Auto-Numerisch	

Bei einigen Modellen kann SQL für das Modellnugget generiert werden, wobei die Modellscoring-Phase per Pushback an die Datenbank zurückgegeben wird. Diese Funktion dient nicht in erster Linie zur Leistungsverbesserung, sondern dazu, dass das gesamte SQL von Streams, die diese Nuggets enthalten, per Pushback zurückgeführt werden kann. Weitere Informationen finden Sie im Thema „Anzeigen von SQL für Modellnuggets“ auf Seite 43.

Tabelle 10. Modellnuggets

SQL-Generierung unterstützendes Modellnugget	Anmerkungen
C&R-Baum	Unterstützung der SQL-Generierung für die Option für einzelne Bäume, nicht jedoch für Boosting, Bagging oder für große Datasetoptionen.
QUEST	
CHAID	

Table 10. Modellnuggets (Forts.)

SQL-Generierung unterstützendes Modellnugget	Anmerkungen
C5.0	
Entscheidungsliste	
Linear	Unterstützung der SQL-Generierung für die Standardmodelloption, nicht jedoch für Boosting, Bagging oder große Datasetoptionen.
Netz	Unterstützung der SQL-Generierung für die Standardmodelloption (nur Mehrschicht-Perzeptron), nicht jedoch für Boosting, Bagging oder große Datasetoptionen.
Faktor/PCA	
Logistisch	Unterstützt SQL-Generierung für das multinomiale, nicht jedoch für das binomiale Verfahren. Beim multinomialen Verfahren wird die Generierung nicht unterstützt, wenn Konfidenzen ausgewählt wurden, es sei denn, der Zieltyp ist "Flag".
Generierte Regelsets	

Table 11. Ausgabe

Knoten zur Unterstützung der SQL-Generierung	Anmerkungen
Tabelle	Unterstützt die Generierung, wenn die SQL-Generierung für den highlight-Ausdruck (siehe Ausdrücke unten) unterstützt wird.
Matrix	Unterstützt die Generierung, außer wenn "Alle numerischen Werte" für die Option "Felder" ausgewählt wurde.
Analyse	Unterstützt die Generierung je nach den ausgewählten Optionen.
Transformieren	
Statistik	Unterstützt die Generierung, wenn die Option "Korrelieren" nicht verwendet wird.
Bericht	
Globalwerte	

Table 12. Exportieren

Knoten zur Unterstützung der SQL-Generierung	Anmerkungen
Datenbank	
Publisher	Der veröffentlichte Stream enthält das generierte SQL.

CLEM-Ausdrücke und Operatoren, die die SQL-Generierung unterstützen

In den folgenden Tabellen erhalten Sie eine Beschreibung der mathematischen Operationen und Ausdrücke, die eine SQL-Generierung unterstützen und die häufig während des Data Minings verwendet werden. Operationen, die in diesen Tabellen nicht enthalten sind, unterstützen die SQL-Generierung in der aktuellen Version nicht.

Tabelle 13. Operatoren

Operation zur Unterstützung der SQL-Generierung	Anmerkungen
+	
-	
/	
*	
><	Wird zur Verkettung von Zeichenfolgen verwendet.

Tabelle 14. Relationale Operatoren

Operation zur Unterstützung der SQL-Generierung	Anmerkungen
=	
/=	Wird verwendet, um "ungleich" anzugeben.
>	
>=	
<	
<=	

Tabelle 15. Funktionen

Operation zur Unterstützung der SQL-Generierung	Anmerkungen
abs	
allbutfirst	
allbutlast	
and	
arccos	
arcsin	
arctan	
arctanh	
cos	
div	
exp	
fracof	
hasstartstring	
hassubstring	
integer	
intof	
isaplhacode	
islowercode	
isnumbercode	
isstartstring	
issubstring	

Table 15. Funktionen (Forts.)

Operation zur Unterstützung der SQL-Generierung	Anmerkungen
isuppercode	
last	
length	
locchar	
log	
log10	
lowertoupper	
max	
member	
min	
negate	
not	
number	
or	
pi	
real	
rem	
round	
sign	
sin	
sqrt	
string	
strmember	
subscrs	
substring	
substring_between	
uppertolower	
to_string	

Table 16. Sonderfunktionen

Operation zur Unterstützung der SQL-Generierung	Anmerkungen
@NULL	
@GLOBAL_AVE	Die speziellen globalen Funktionen werden zum Abrufen von Globalwerten verwendet, die vom Globalwerteknoten berechnet werden.
@GLOBAL_SUM	
@GLOBAL_MAX	
@GLOBAL_MEAN	
@GLOBAL_MIN	
@GLOBALSDEV	

Tabelle 17. Aggregierungsfunktionen

Operation zur Unterstützung der SQL-Generierung	Anmerkungen
Summe	
Mittelwert	
Min	
Max	
Anzahl	
Standardabweichung	

Verwenden von SQL-Funktionen in CLEM-Ausdrücken

Mit der Funktion @SQLFN können benannte SQL-Funktionen innerhalb von CLEM-Ausdrücken zum abschließlichen Zweck der Datenbankausführung hinzugefügt werden. Diese Vorgehensweise kann in bestimmten Fällen sinnvoll sein, wenn proprietäres SQL oder andere herstellerspezifische Anpassungen erforderlich sind.

Die Verwendung dieser Funktion wird nicht durch die standardmäßige IBM SPSS Modeler-Supportvereinbarung abgedeckt, da die Ausführung von externen Datenbankkomponenten abhängt, die sich der Kontrolle von IBM entziehen. Eine Bereitstellung ist jedoch in Sonderfällen möglich, normalerweise im Rahmen einer Dienstleistungsabgabe. Weitere Informationen erhalten Sie, falls erforderlich, unter <http://www.ibm.com/software/analytics/spss/services/>.

Schreiben von SQL-Abfragen

Wenn Sie den Datenbankknoten verwenden, sollten Sie besonders auf alle SQL-Abfragen achten, die zu einem Dataset mit doppelten Spaltennamen führen. Diese doppelten Namen verhindern häufig die SQL-Optimierung für alle nachgeordneten Knoten.

IBM SPSS Modeler verwendet verschachtelte SELECT-Anweisungen, um ein Pushback von SQL für Streams durchzuführen, die eine SQL-Abfrage im Datenbankquellenknoten verwenden. Mit anderen Worten: Der Stream verschachtelt die im Datenbankquellenknoten angegebene Abfrage in eine oder mehrere SELECT-Anweisungen, die während der Optimierung von nachgeordneten Knoten erstellt werden. Wenn also das Ergebnisset einer Abfrage doppelte Spaltennamen enthält, kann die Anweisung nicht vom RDBMS verschachtelt werden. Verschachtelungsschwierigkeiten treten häufig während einer Tabellenzusammenfügung auf, bei der eine Spalte mit demselben Namen in mehr als einer der zusammengeführten Tabellen ausgewählt ist. Beachten Sie beispielsweise folgende Abfrage im Quellenknoten:

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

Die Abfrage verhindert eine nachfolgende SQL-Optimierung, da diese SELECT-Anweisung zu einem Dataset mit zwei Spalten mit der Bezeichnung ID führen würde.

Um eine vollständige SQL-Optimierung zu ermöglichen, sollten Sie beim Schreiben von SQL-Abfragen genau sein und die Spaltenaliasnamen angeben, wenn eine Situation mit doppelten Spaltennamen auftritt. Die Anweisung unten zeigt eine genauere Abfrage:

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

Anhang A. Konfigurieren von Oracle für UNIX-Plattformen

Konfigurieren von Oracle für die SQL-Optimierung

Wenn Sie IBM SPSS Modeler Server auf UNIX-Plattformen ausführen und die Daten aus einer Oracle-Datenbank gelesen werden, sollten Sie die folgenden Tipps beachten, um sicherzustellen, dass das generierte SQL in der Datenbank vollständig optimiert wird.

Richtige Ländereinstellung

Wenn Sie IBM SPSS Modeler Server mit einer Ländereinstellung ausführen, die von der mit den Connect-ODBC-Treibern mitgelieferten abweicht, sollten Sie den Computer neu konfigurieren, um so die SQL-Optimierung zu verbessern. Die Connect-ODBC-Treiber werden lediglich mit den Dateien für die Ländereinstellung *en_US* geliefert. Wenn der IBM SPSS Modeler Server-Computer mit einer anderen Ländereinstellung ausgeführt wird oder wenn für die Shell, in der IBM SPSS Modeler Server gestartet wurde, die Ländereinstellung nicht vollständig definiert wurde, wird das generierte SQL u. U. in Oracle nicht vollständig optimiert. Dafür gibt es folgende Gründe:

- IBM SPSS Modeler Server verwendet die Dateien für die ODBC-Ländereinstellungen entsprechend der Ländereinstellung, mit der sie ausgeführt werden, um die von der Datenbank in Textzeichenfolgen ausgegebenen Codes zu übersetzen. Diese Textzeichenfolgen werden dann verwendet, um zu ermitteln, mit welcher Datenbank eine Verbindung hergestellt wird.
- Wenn die Ländereinstellung (wie von der Systemabfrage `$LANG` an IBM SPSS Modeler Server ausgegeben) nicht *en_US* lautet, kann IBM SPSS Modeler die Codes, die es vom ODBC-Treiber erhält, nicht in Text übersetzen. Mit anderen Worten: Zu Beginn der Datenbankverbindung wird anstelle der Zeichenfolge *Oracle* ein unübersetzter Code an IBM SPSS Modeler Server ausgegeben. Dies bedeutet, dass IBM SPSS Modeler keine Streams für Oracle optimieren kann.

So prüfen Sie die Ländereinstellungen und setzen Sie zurück:

1. Führen Sie in einer UNIX-Shell folgenden Befehl aus:

```
#locale
```

Sie erhalten Informationen über die Ländereinstellung für die Shell. Beispiel:

```
$ 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. Wechseln Sie in das Connect-Verzeichnis ODBC/locale. (Hier sehen Sie das Unterverzeichnis *en_US*.)
3. Erstellen Sie eine Verknüpfung zu diesem Verzeichnis *en_US*, indem Sie den Namen der Ländereinstellung der Shell angeben. Hier ein Beispiel:

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

Für eine anderssprachige Ländereinstellung, wie z. B. *fr_FR.ISO8859-1*, sollten Sie die Verknüpfung wie folgt herstellen:

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

4. Nachdem Sie den Link erstellt haben, müssen Sie IBM SPSS Modeler Server aus derselben Shell neu starten. (IBM SPSS Modeler Server erhält die jeweiligen Ländereinstellungsinformationen aus der Shell, aus der das Produkt gestartet wurde.)

Anmerkungen

Wenn Sie ein UNIX-System für SQL-Pushbacks nach Oracle optimieren, sollten Sie folgende Tipps beachten:

- Die vollständige Ländereinstellung muss angegeben werden. Im obigen Beispiel müssen Sie die Verknüpfung in der Form `language_territory.code-page` erstellen. Das bereits vorhandene Verzeichnis für die Ländereinstellung `en_US` reicht nicht aus.
- Um das datenbankinterne Mining zu optimieren, müssen sowohl `LANG` als auch `LC_ALL` in der zum Starten von IBM SPSS Modeler Server verwendeten Shell definiert sein. `LANG` kann in der Shell genauso definiert werden wie jede andere Umgebungsvariable, bevor IBM SPSS Modeler Server neu gestartet wird. Siehe folgende Definition:

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

- Bei jedem Start von IBM SPSS Modeler Server müssen Sie prüfen, ob die Shell-Informationen zur Ländereinstellung vollständig angegeben sind und ob die entsprechende Verknüpfung im Verzeichnis `ODBC/locale` vorhanden ist.

Anhang B. Konfigurieren von UNIX-Startscripts

Einführung

In diesem Anhang werden einige der Scripts beschrieben, die im Lieferumfang der UNIX-Versionen von IBM SPSS Modeler Server enthalten sind. Außerdem erfahren Sie hier, wie Sie die Scripts konfigurieren können. Scripts werden für die folgenden Zwecke verwendet:

- Konfigurieren von IBM SPSS Modeler Server für einen automatischen Start beim Neustart des Server-Computers.
- Manuelles Stoppen und Neustarten von IBM SPSS Modeler Server.
- Ändern der Berechtigungen für die von IBM SPSS Modeler Server erstellten Dateien.
- Konfigurieren von IBM SPSS Modeler Server, um mit den in IBM SPSS Modeler Server enthaltenen ODBC Connect-Treibern zu arbeiten. Weitere Informationen finden Sie im Thema „IBM SPSS Modeler Server und das Data Access Pack“ auf Seite 56.
- Konfigurieren von IBM SPSS Modeler Server für die Verwendung des IBM SPSS Collaboration and Deployment Services Enterprise View Driver-Treibers beim Lesen von Daten mit einem Enterprise View-Knoten. Weitere Informationen finden Sie im Thema „Konfigurieren eines Treibers für den Enterprise-Ansichtsknoten“ auf Seite 62.

Scripts

IBM SPSS Modeler Server verwendet verschiedene Scripts, darunter:

- **modelersrv.sh.** Das Script für den manuellen Start von IBM SPSS Modeler Server befindet sich im Installationsverzeichnis von IBM SPSS Modeler Server. Es konfiguriert die Serverumgebung, wenn der Serverdämonprozess *manuell* gestartet wird. Führen Sie es aus, wenn Sie den Server manuell starten und stoppen möchten. Bearbeiten Sie es, wenn Sie die Konfiguration für den manuellen Start ändern möchten.
- **auto.sh.** Mit diesem Script wird Ihr System konfiguriert, um den Serverdämonprozess beim Neustart automatisch zu starten. Führen Sie das Script ein Mal aus, um Ihr System für den automatischen Start zu konfigurieren. Sie müssen es nicht bearbeiten. Das Script befindet sich im Installationsverzeichnis von IBM SPSS Modeler Server.
- **rc.modeler.** Beim Ausführen von *auto.sh* wird dieses Script an einem Speicherort erstellt, der vom Betriebssystem Ihres Servers abhängt. Es konfiguriert die Serverumgebung, wenn der Server *automatisch* gestartet wird. Bearbeiten Sie es, wenn Sie die Konfiguration für den automatischen Start ändern möchten.

Tabelle 18. Speicherort von "rc.modeler" unter den verschiedenen Betriebssystemen

Betriebssystem	Ort
AIX	/etc/rc.modeler
HP-UX	/sbin/init.d/rc.modeler
Solaris	/etc/init.d/rc.modeler

Automatisches Starten und Stoppen von IBM SPSS Modeler Server

IBM SPSS Modeler Server muss als Dämonprozess gestartet werden. Im Installationsprogramm ist das Script *auto.sh* enthalten, mit dem Sie Ihr System für das automatische Stoppen und Neustarten von IBM SPSS Modeler Server konfigurieren können.

So konfigurieren Sie das System für das automatische Starten und Herunterfahren:

1. Melden Sie sich als Root an.
2. Wechseln Sie zum IBM SPSS Modeler Server-Installationsverzeichnis.
3. Führen Sie das Script aus. Geben Sie an der UNIX-Eingabeaufforderung Folgendes ein:
`./auto.sh`

Das automatische Startscript *rc.modeler* wird an dem in der vorangehenden Tabelle angegebenen Speicherort erstellt. Das Betriebssystem verwendet *rc.modeler*, um den IBM SPSS Modeler Server-Dämonprozess zu starten, wenn der Server-Computer neu gestartet wird. Außerdem verwendet das Betriebssystem *rc.modeler* auch, um den Dämon zu stoppen, wenn der Computer heruntergefahren wird.

Manuelles Starten und Stoppen von IBM SPSS Modeler Server

Sie können IBM SPSS Modeler Server manuell starten und stoppen, indem Sie das Script *modelersrv.sh* ausführen.

So können Sie IBM SPSS Modeler Server manuell starten und stoppen:

1. Wechseln Sie zum IBM SPSS Modeler Server-Installationsverzeichnis.
2. Um den Server zu starten, geben Sie an der UNIX-Eingabeaufforderung Folgendes ein:
`./modelersrv.sh start`
3. Um den Server anzuhalten, geben Sie an der UNIX-Eingabeaufforderung Folgendes ein:
`./modelersrv.sh stop`

Bearbeiten von Scripts

Wenn Sie sowohl den manuellen als auch den automatischen Start verwenden, nehmen Sie die Änderungen parallel in *modelersrv.sh* und *rc.modeler* vor. Wenn Sie nur den manuellen Start verwenden, bearbeiten Sie nur *modelersrv.sh*. Wenn Sie nur den automatischen Start verwenden, bearbeiten Sie nur *rc.modeler*.

So bearbeiten Sie Scripts:

1. Stoppen Sie IBM SPSS Modeler Server. (Weitere Informationen finden Sie im Thema „Manuelles Starten und Stoppen von IBM SPSS Modeler Server“.)
2. Wählen Sie das entsprechende Script aus. (Weitere Informationen finden Sie im Thema „Scripts“ auf Seite 55.)
3. Öffnen Sie das Script in einem Texteditor, nehmen Sie die Änderungen vor und speichern Sie die Datei.
4. Starten Sie IBM SPSS Modeler Server, entweder automatisch (indem Sie den Server-Computer neu starten) oder manuell.

Steuern der Berechtigungen für die Dateierstellung

IBM SPSS Modeler Server erstellt temporäre Dateien mit Lese-, Schreib- und Ausführungsberechtigungen für alle Benutzer. Sie können diese Standardeinstellung überschreiben, indem Sie im Startscript *modelersrv.sh*, *rc.modeler* oder in beiden die Einstellung *UMASK* bearbeiten. (Weitere Informationen finden Sie unter „Bearbeiten von Scripts“.) Wir empfehlen als restriktivste *UMASK*-Einstellung 077. Restriktivere Einstellungen könnten zu Berechtigungsproblemen für IBM SPSS Modeler Server führen.

IBM SPSS Modeler Server und das Data Access Pack

Wenn Sie die ODBC-Treiber mit IBM SPSS Modeler Server verwenden möchten, muss die ODBC-Umgebung mit *odbc.sh* konfiguriert werden, wenn der IBM SPSS Modeler Server-Prozess startet. Sie erreichen dies, indem Sie das entsprechende IBM SPSS Modeler-Startscript in *modelersrv.sh*, *rc.modeler* oder in beiden konfigurieren. (Weitere Informationen finden Sie im Thema „Bearbeiten von Scripts“.)

Weitere Informationen finden Sie auf der Website des Technical Support unter <http://www.ibm.com/support>. Wenn Sie Fragen zur Erstellung oder Einstellung von Berechtigungen für ODBC-Datenquellen haben, wenden Sie sich an Ihren Datenbankadministrator.

So konfigurieren Sie ODBC für den Start mit IBM SPSS Modeler Server:

1. Stoppen Sie den IBM SPSS Modeler Server-Host, falls er gerade ausgeführt wird.
2. Laden Sie das relevante komprimierte TAR-Archiv für die Plattform herunter, auf der IBM SPSS Modeler Server installiert ist. Stellen Sie sicher, dass Sie die Treiber für die korrekte Version von IBM SPSS Modeler Server, die Sie installiert haben, herunterladen. Kopieren Sie die Datei an den Speicherort, an dem Sie die ODBC-Treiber installieren möchten (z. B. */usr/spss/odbc*).
3. Extrahieren Sie das Archiv.
4. Führen Sie das *setodbcpath.sh*-Script aus, das aus dem Archiv extrahiert wird.
5. Bearbeiten Sie das Script *odbc.sh*, um die Definition von ODBCINI unten zu diesem Script hinzuzufügen und sie zu exportieren, z. B.:

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

ODBCINI muss auf den vollständigen Pfadnamen der *odbc.ini*-Datei verweisen, die von IBM SPSS Modeler gelesen werden soll, um eine Liste der ODBC-Datenquellen, die Sie definieren, zu erhalten (eine Standard-odbc.ini wurde mit den Treibern installiert).

6. Speichern Sie *odbc.sh*.
7. (Nur 64-Bit-IBM SPSS Modeler Server-Installationen, fahren Sie für alle anderen Installationen beim nächsten Schritt fort) Definieren und exportieren Sie LD_LIBRARY_PATH_64 in *odbc.sh*:

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

Dabei ist *library_path* identisch mit dem Wert der LD_LIBRARY_PATH-Definition, die bereits in dem Script enthalten ist, das mit Ihrem Installationspfad initialisiert wurde (z. B. */usr/spss/odbc/lib*). Der einfachste Weg ist, die Anweisungen *if* und *export* für LD_LIBRARY_PATH in Ihre *odbc.sh*-Datei zu kopieren, sie an das Ende der Datei anzuhängen und dann die LD_LIBRARY_PATH-Zeichenfolgen in den neu angehängten *if*- und *export*-Anweisungen durch LD_LIBRARY_PATH_64 zu ersetzen.

Ihre finale *odbc.sh*-Datei auf einer 64-Bit-IBM SPSS Modeler Server-Installationen könnte also wie folgt aussehen:

```
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
```

Denken Sie daran, LD_LIBRARY_PATH_64 zu exportieren und mit der *if*-Schleife zu definieren.

8. Bearbeiten Sie die *odbc.ini*-Datei, die Sie vorher mit \$ODBCINI definiert haben. Definieren Sie die Datenquellennamen, die Sie benötigen (diese hängen von der Datenbank ab, auf die sie zugreifen).
9. Speichern Sie die *odbc.ini*-Datei.

10. Konfigurieren Sie IBM SPSS Modeler Server, um diese Treiber zu verwenden. Bearbeiten Sie dazu *modelersrv.sh* und fügen Sie die folgende Zeile direkt unter die Zeile ein, die SCLEMDNAME definiert:

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

Dabei ist *odbc.sh_path* der vollständige Pfad zur Datei *odbc.sh*, die Sie zu Beginn dieses Vorgangs bearbeitet haben, z. B.:

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

Hinweis: Hier ist die Syntax wichtig; achten Sie darauf, dass zwischen dem ersten Punkt und dem Pfad zu Datei ein Leerzeichen ist.

11. Speichern Sie *modelersrv.sh*.

So testen Sie die Verbindung:

1. Starten Sie IBM SPSS Modeler Server neu.
2. Verbinden Sie sich von einem Client mit IBM SPSS Modeler Server.
3. Fügen Sie auf dem Client einen Datenbankquellenknoten zum Erstellungsbereich hinzu.
4. Öffnen Sie den Knoten und stellen Sie sicher, dass Sie die Datenquellennamen sehen können, die Sie zu einem früheren Zeitpunkt des Konfigurationsvorgangs in der *odbc.ini*-Datei definiert haben.

Wenn die Namen nicht angezeigt werden oder Sie Fehlermeldungen bei dem Versuch, sich mit einer von Ihnen definierten Datenquelle zu verbinden, erhalten, folgen Sie dem Verfahren zur Fehlerbehebung. Weitere Informationen finden Sie im Thema „Fehlerbehebung bei der ODBC-Konfiguration“ auf Seite 59.

So konfigurieren Sie ODBC für den Start mit IBM SPSS Modeler Solution Publisher Runtime:

Wenn Sie erfolgreich von IBM SPSS Modeler Server eine Verbindung zur Datenbank aufbauen können, können Sie eine IBM SPSS Modeler Solution Publisher Runtime-Installation auf dem gleichen Server konfigurieren, indem Sie das gleiche *odbc.sh*-Script aus dem Startscript von IBM SPSS Modeler Solution Publisher Runtime referenzieren.

1. Bearbeiten Sie das *modelerrun*-Script in IBM SPSS Modeler Solution Publisher Runtime, um die folgende Zeile direkt unterhalb der letzten Zeile des Scripts einzufügen:

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

Dabei ist *odbc.sh_path* der vollständige Pfad zur Datei *odbc.sh*, die Sie zu Beginn dieses Vorgangs bearbeitet haben, z. B.:

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

Hinweis: Hier ist die Syntax wichtig; achten Sie darauf, dass zwischen dem ersten Punkt und dem Pfad zu Datei ein Leerzeichen ist.

2. Speichern Sie die *modelerrun*-Scriptdatei.

So konfigurieren Sie ODBC für den Start mit IBM SPSS Modeler Batch:

Für ODBC ist keine Konfiguration des IBM SPSS Modeler Batch-Scripts erforderlich. Das liegt daran, dass Sie sich mit IBM SPSS Modeler Server von IBM SPSS Modeler Batch verbinden, um Streams auszuführen. Stellen Sie sicher, dass die IBM SPSS Modeler Server-ODBC-Konfiguration durchgeführt wurde und ordnungsgemäß funktioniert, wie weiter oben in diesem Abschnitt beschrieben.

So fügen Sie einen Datenquellennamen hinzu oder bearbeiten ihn:

1. Bearbeiten Sie die *odbc.ini*-Datei, um den neuen oder geänderten Namen aufzunehmen.
2. Testen Sie die Verbindung wie weiter oben in diesem Abschnitt beschrieben.

Wenn die Verbindung mit IBM SPSS Modeler Server ordnungsmäßig funktioniert, sollte die neue oder geänderte Datenquelle auch korrekt mit IBM SPSS Modeler Solution Publisher Runtime und IBM SPSS Modeler Batch funktionieren.

Fehlerbehebung bei der ODBC-Konfiguration

Keine Datenquellen aufgelistet oder zufälliger Text angezeigt

Wenn Sie einen Datenbankquellenknoten öffnen und die Liste der verfügbaren Datenquellen leer ist oder unerwartete Einträge enthält, kann dies an einem Problem mit dem Startscript liegen.

1. Überprüfen Sie, ob `$ODBCINI` in `modelersrv.sh` definiert ist, entweder ausdrücklich im Script selbst oder im `odbc.sh`-Script, das in `modelersrv.sh` referenziert ist.
2. Stellen Sie in letzterem Fall sicher, dass `ODBCINI` auf den vollständigen Pfad zur `odbc.ini`-Datei verweist, die Sie verwendet haben, um Ihre ODBC-Datenquellen zu definieren.
3. Wenn die Pfadangabe in `ODBCINI` richtig ist, überprüfen Sie den `$ODBCINI`-Wert, der in der IBM SPSS Modeler Server-Umgebung verwendet wird, indem Sie die Variable aus `modelersrv.sh` wiederholen. Fügen Sie dazu die folgende Zeile zu `modelersrv.sh` hinzu, und zwar nach der Position, an der Sie `ODBCINI` definieren:

```
echo $ODBCINI
```

4. Speichern Sie und führen Sie `modelersrv.sh` dann aus. Der in der IBM SPSS Modeler Server-Umgebung festgelegte `$ODBCINI`-Wert wird zur Verifizierung in die Standardausgabe geschrieben.
5. Wenn kein Wert an die Standardausgabe zurückgegeben wird und Sie `$ODBCINI` im `odbc.sh`-Script, das Sie von `modelersrv.sh` referenzieren, definieren, überprüfen Sie, ob die Referenzierungssyntax korrekt ist. Sie sollte wie folgt aussehen:

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

Dabei ist `odbc.sh_path` der vollständige Pfad zur Datei `odbc.sh`, die Sie zu Beginn dieses Vorgangs bearbeitet haben, z. B.:

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

Hinweis: Hier ist die Syntax wichtig; achten Sie darauf, dass zwischen dem ersten Punkt und dem Pfad zu Datei ein Leerzeichen ist.

Wenn der korrekte Wert bei ausgeführter `modelersrv.sh` in der Standardausgabe wiederholt wird, sollten Sie die Datenquellennamen im Datenbankquellenknoten sehen können, wenn Sie IBM SPSS Modeler Server neu starten und vom Client eine Verbindung herstellen.

Der IBM SPSS Modeler-Client hängt beim Klicken auf das Dialogfeld "In Datenbankverbindungen verbinden"

Dieses Verhalten kann von Ihrem Bibliothekspfad verursacht werden, der nicht korrekt festgelegt wurde, sodass er den Pfad zu den ODBC-Bibliotheken enthält. Der Bibliothekspfad wird definiert von `$LD_LIBRARY_PATH` (und `$LD_LIBRARY_PATH_64` bei 64-Bit-Versionen).

Um den Wert des Bibliothekspfads in der IBM SPSS Modeler Server-Dämonumgebung anzuzeigen, wiederholen Sie den Wert der entsprechenden Umgebungsvariable aus `modelersrv.sh` nach der Zeile, in der Sie den ODBC-Bibliothekspfad an den Bibliothekspfad anhängen, und führen Sie das Script aus. Der Wert des Bibliothekspfads wird im Terminal wiederholt, wenn Sie das Script das nächste Mal ausführen.

Wenn Sie `odbc.sh` aus `modelersrv.sh` referenzieren, um Ihre IBM SPSS Modeler Server-ODBC-Umgebung einzurichten, wiederholen Sie den Bibliothekspfadwert aus der Zeile nach der, in der Sie das `odbc.sh`-Script referenzieren. Um den Wert zu wiederholen, fügen Sie die folgende Zeile zum Script hinzu, speichern Sie dann und führen Sie die Scriptdatei aus:

```
echo <library_path_variable>
```

Dabei ist `<library_path_variable>` die geeignete Bibliothekspfadvariable für Ihr Serverbetriebssystem.

Der zurückgegebene Wert Ihres Bibliothekspfads muss den Pfad zum `lib`-Unterverzeichnis Ihrer ODBC-Installation enthalten. Wenn dies nicht der Fall ist, hängen Sie diesen Speicherort an die Datei an.

Wenn Sie die 64-Bit-Version von IBM SPSS Modeler Server ausführen, überschreibt \$LD_LIBRARY_PATH_64 \$LD_LIBRARY_PATH, wenn diese Variable festgelegt ist. Wenn dieses Problem bei einer der 64-Bit-Plattformen auftritt, wiederholen Sie LD_LIBRARY_PATH_64 und \$LD_LIBRARY_PATH aus *modelersrv.sh* und, falls erforderlich, legen Sie \$LD_LIBRARY_PATH_64 so fest, dass diese Variable das *lib*-Unterverzeichnis Ihrer ODBC-Installation enthält, und exportieren Sie die Definition.

Datenquellename nicht gefunden und kein Standardtreiber angegeben

Wenn dieser Fehler beim Klicken auf das Dialogfeld "In Datenbankverbindungen verbinden" angezeigt wird, bedeutet dies meist, dass Ihre *odbc.ini* nicht richtig definiert wurde. Stellen Sie sicher, dass der Datenquellename (DSN), wie er im Abschnitt [ODBC Data Sources] oben in der Datei definiert ist, mit der Zeichenfolge zwischen den eckigen Klammern weiter unten in *odbc.ini* übereinstimmt, um den DSN zu definieren. Wenn diese sich in irgendeiner Art und Weise unterscheiden, wird diese Fehlernachricht angezeigt, wenn Sie eine Verbindung unter Verwendung des DSN von IBM SPSS Modeler herstellen möchten. Es folgt ein Beispiel einer *falschen* Spezifikation:

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

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

Sie müssen eine der beiden fett markierten Zeichenfolgen ändern, damit sie genau übereinstimmen. Dies sollte den Fehler beheben.

Angegebener Treiber konnte nicht geladen werden

Dieser Fehler gibt auch an, dass die *odbc.ini*-Datei falsch definiert ist. Eine Möglichkeit ist, dass der Treiberparameter in der Treiberstanza falsch festgelegt ist, z. B.:

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

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

1. Überprüfen Sie, ob das vom Treiberparameter angegebene gemeinsam genutzte Objekt vorhanden ist.
2. Korrigieren Sie den Pfad zum gemeinsam genutzten Objekt, falls dieser falsch ist.
3. Wenn der Treiberparameter im Format

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

spezifiziert ist, bedeutet dies, dass Sie Ihre ODBC-bezogenen Scripts nicht initialisiert haben. Führen Sie das *setodbcpath.sh*-Script aus, das mit den Treibern installiert wurde. Weitere Informationen finden Sie im Thema „IBM SPSS Modeler Server und das Data Access Pack“ auf Seite 56. Wenn Sie dieses Script ausgeführt haben, sollten Sie sehen, dass die Zeichenfolge "ODBCHOME" durch den Pfad zu Ihrer ODBC-Installation ersetzt wurde. Dies sollte das Problem lösen.

Ein weiterer Grund könnte ein Problem mit der Bibliothek des Treibers sein. Verwenden Sie das in ODBC enthaltene Tool `ivtestlib`, um zu bestätigen, dass der Treiber nicht geladen werden kann. Bei Connect64 verwenden Sie das Tool `ddtestlib`. Beheben Sie das Problem, indem Sie die Variable für den Bibliothekspfad im Startscript festlegen.

Wenn beispielsweise der Oracle-Treiber für eine 32-Bit-Installation nicht geladen werden kann, führen Sie folgende Schritte durch:

1. Verwenden Sie `ivtestlib`, um zu bestätigen, dass der Treiber nicht geladen werden kann. Geben Sie beispielsweise an der UNIX-Eingabeaufforderung Folgendes ein:

```
sh
cd ODBC DIR
. odbcd.sh
./bin/ivtestlib MFor815
```

Hierbei wird `ODBC DIR` durch den Pfad zum ODBC-Installationsverzeichnis ersetzt.

2. Lesen die Sie Nachricht, um festzustellen, ob ein Fehler vorliegt. Beispiel:

Load of MFor815.so failed: ld.so.1: bin/ivtestlib: fatal: libclntsh.so: open failed: No such file or directory

Diese Nachricht gibt an, dass die Oracle-Client-Bibliothek `libclntsh.so` nicht vorhanden ist oder sich nicht im angegebenen Pfad befindet (z. B. nicht unter `LD_LIBRARY_PATH` unter Solaris).

3. Stellen Sie fest, ob die Bibliothek vorhanden ist. Falls nicht, installieren Sie den Oracle-Client erneut. Ist die Bibliothek vorhanden, geben Sie an der UNIX-Eingabeaufforderung die folgende Befehlssequenz ein:

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

Hierbei wird `/bigdisk/oracle/product/8.1.6/lib` durch den Pfad `libclntsh.so` ersetzt. `LD_LIBRARY_PATH` ist die Variable für den Bibliothekspfad für Ihr Betriebssystem.

Beachten Sie: Wenn Sie die 64-Bit-Version von IBM SPSS Modeler unter Linux oder Solaris ausführen, enthält die Variable für den Bibliothekspfad das Suffix `_64`. Die ersten beiden Zeilen aus dem vorstehenden Beispiel lauten hier also entsprechend:

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

4. Lesen Sie die Nachricht, um festzustellen, ob der Treiber nun geladen werden kann. Beispiel:

Load of MFor815.so successful, qehandle is 0xFF3A1BE4

Diese Nachricht gibt an, dass die Oracle-Client-Bibliothek geladen werden kann.

5. Berichtigen Sie den Bibliothekspfad im Startscript von IBM SPSS Modeler.
6. Starten Sie IBM SPSS Modeler Server mit dem Startscript neu, das Sie bearbeitet haben (`modelersrv.sh` bzw. `rc.modeler`).

Bibliothekspfade

Der Name der Variablen für den Bibliothekspfad hängt von Ihrem Betriebssystem ab. Die folgende Tabelle dient als Hilfestellung bei den entsprechenden Einstellungen, wenn Sie Ihr System konfigurieren oder Fehler beheben.

Tabelle 19. Bibliothekspfade (nach Betriebssystem geordnet)

Betriebssystem	Name der Bibliothekspfadvariablen
AIX	LIBPATH
HP-UX	SHLIB_PATH
Solaris oder Linux (32-Bit)	LD_LIBRARY_PATH
Solaris oder Linux (64-Bit)	LD_LIBRARY_PATH_64

Konfigurieren eines Treibers für den Enterprise-Ansichtsknoten

Mit dem Enterprise-Ansichtsknoten können Sie eine Verbindung zwischen einer IBM SPSS Modeler-Sitzung und einer IBM SPSS Collaboration and Deployment Services Enterprise View in einem freigegebenen IBM SPSS Collaboration and Deployment Services Repository herstellen und aufrechterhalten.

Um den Enterprise-Ansichtsknoten mit IBM SPSS Modeler Server verwenden zu können, muss ein IBM SPSS Collaboration and Deployment Services Enterprise View Driver installiert und konfiguriert sein. Details zur Installation des Treibers erhalten Sie bei Ihrem lokalen Administrator. Nach der Installation muss ein Verweis auf *pev.sh* in das Startscript aufgenommen werden - entweder in *modelersrv.sh*, in *rc.modeler* oder in beiden -, das folgendes Format aufweist:

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

Anhang C. Konfigurieren und Ausführen von SPSS Modeler Server als Nicht-Root-Prozess unter UNIX

Einführung

In den folgenden Anweisungen finden Sie Informationen zur Ausführung von IBM SPSS Modeler Server als Nicht-Root-Prozess unter UNIX.

Ausführen als Root. Die Standardinstallation von IBM SPSS Modeler Server geht davon aus, dass der Serverdämonprozess als Root ausgeführt wird. Bei der Ausführung als Root kann IBM SPSS Modeler jede Benutzeranmeldung zuverlässig authentifizieren und jede Benutzersitzung mit dem entsprechenden UNIX-Benutzerkonto starten. Auf diese Weise wird sichergestellt, dass Benutzer ausschließlich Zugriff auf ihre eigenen Dateien und Verzeichnisse erhalten.

Ausführen als Nicht-Root. Um IBM SPSS Modeler Server als Nicht-Root-Prozess auszuführen, werden die echten und effektiven Benutzer-IDs des Serverdämonprozesses auf ein Konto Ihrer Wahl gesetzt. Alle von SPSS Modeler Server gestarteten Benutzersitzungen verwenden dasselbe UNIX-Konto. Dies bedeutet, dass alle Dateidaten, die von SPSS Modeler gelesen oder geschrieben werden, für alle SPSS Modeler-Benutzer freigegeben werden. Dies gilt nicht für den Zugriff auf Datenbankdaten, da sich Benutzer unabhängig hiervon bei allen den von ihnen verwendeten Datenbank-Datenquellen authentifizieren müssen. Ohne Rootberechtigung funktioniert IBM SPSS Modeler auf eine der beiden folgenden Weisen:

- **Ohne private Kennwortdatenbank.** Bei dieser Methode verwendet SPSS Modeler die bestehende UNIX-Kennwortdatenbank, NIS oder den LDAP-Server, der normalerweise für die Authentifizierung auf dem UNIX-System verwendet wird. Weitere Informationen finden Sie im Thema „Konfigurieren als Nicht-Root-Benutzer ohne private Kennwortdatenbank“.
- **Mit einer privaten Kennwortdatenbank.** Bei dieser Methode authentifiziert SPSS Modeler die Benutzer anhand einer privaten Kennwortdatenbank, die von der bestehenden UNIX-Kennwortdatenbank, NIS bzw. dem LDAP-Server, der normalerweise für die Authentifizierung auf dem UNIX-System verwendet wird, getrennt ist. Weitere Informationen finden Sie im Thema „Konfigurieren als Nicht-Root-Benutzer mit einer privaten Kennwortdatenbank“ auf Seite 64.

Konfigurieren als Nicht-Root-Benutzer ohne private Kennwortdatenbank

So konfigurieren Sie IBM SPSS Modeler Server für die Ausführung mit einem Nicht-Root-Konto, ohne dass eine private Kennwortdatenbank erforderlich ist:

1. Öffnen Sie die SPSS Modeler Server-Datei *options.cfg* zum Bearbeiten.
2. Setzen Sie die Option `start_process_as_login_user` auf Y.
3. Speichern und schließen Sie die Datei *options.cfg*.

Bei der Ausführung als Nicht-Root-Benutzer sind wahrscheinlich Konfigurationsaktualisierungen erforderlich. Weitere Informationen finden Sie im Thema „Fehlerbehebung bei fehlgeschlagener Benutzerauthentifizierung“ auf Seite 65.

Vorsicht: Sie sollten nicht die Einstellung `start_process_as_login_user` aktivieren und dann IBM SPSS Modeler Server als *Root* starten. Dadurch würden für alle Benutzer, die mit dem Server verbunden sind, die Serverprozesse als *Root* ausgeführt, was ein Sicherheitsrisiko darstellen würde.

Konfigurieren als Nicht-Root-Benutzer mit einer privaten Kennwortdatenbank

Wenn Sie auswählen, dass die Benutzer mittels einer privaten Kennwortdatenbank authentifiziert werden sollen, werden alle Sitzungen mit demselben Nicht-Root-Benutzerkonto gestartet.

So konfigurieren Sie IBM SPSS Modeler Server für eine derartige Ausführung als Nicht-Root-Konto:

1. Erstellen Sie eine Gruppe, die alle Benutzer enthält. Sie können dieser Gruppe einen beliebigen Namen geben. Für dieses Beispiel wird *modelerusers* verwendet.
2. Erstellen Sie das Benutzerkonto, auf dem IBM SPSS Modeler Server ausgeführt werden soll. Dieses Konto dient zur ausschließlichen Verwendung für den IBM SPSS Modeler Server-Dämonprozess. Für dieses Beispiel wird der Name *modelerserv* verwendet.

Beachten Sie beim Erstellen des Kontos Folgendes:

- Die Hauptgruppe sollte die soeben erstellte Gruppe *<modelerusers>* sein.
 - Das Basisverzeichnis kann das Installationsverzeichnis von IBM SPSS Modeler sein oder jedes andere geeignete Standardverzeichnis (wenn das Konto auch nach Aufrüstungen erhalten bleiben soll, sollten Sie nicht das Installationsverzeichnis verwenden).
3. Konfigurieren Sie anschließend die Startscripts, um IBM SPSS Modeler Server mit dem neu erstellten Konto zu starten. Öffnen Sie das entsprechende Startscript in einem Texteditor. Weitere Informationen finden Sie im Thema „Scripts“ auf Seite 55.

- a. Ändern Sie die Einstellung *umask*, um mindestens Gruppenlesezugriff für erstellte Dateien zu erteilen:

```
umask 027
```

4. Bearbeiten Sie die Datei mit den Serveroptionen *config/options.cfg*, um die Authentifizierung mit der privaten Kennwortdatenbank einzustellen, indem Sie die folgende Zeile anhängen:

```
authentication_methods, "pasw_modeler"
```

5. Setzen Sie die Option *start_process_as_login_user* auf *Y*.

6. Anschließend erstellen Sie eine private Kennwortdatenbank, die in der Datei *config/passwords.cfg* gespeichert wird. Die Kennwortdatei definiert die Kombinationen aus Benutzername/Kennwort, die für die Anmeldung bei IBM SPSS Modeler zulässig sind. *Hinweis:* Diese Daten sind nur für IBM SPSS Modeler bestimmt (privat) und stehen nicht in Zusammenhang mit den Benutzernamen und Kennwörtern, die für die Anmeldung an UNIX verwendet werden. Sie können dieselben Benutzernamen verwenden, wenn dies praktikabel ist, die Verwendung derselben Kennwörter ist jedoch nicht zulässig.

Zum Erstellen der Kennwortdatei benötigen Sie das Dienstprogramm für Kennwörter (*pwutil*), das sich im Verzeichnis *bin* der IBM SPSS Modeler Server-Installation befindet. Die Synopse dieses Programms lautet:

```
pwutil [ Benutzername [ Kennwort ] ]
```

Das Programm nimmt einen Benutzernamen und ein Nur-Text-Kennwort und schreibt den Benutzernamen und das verschlüsselte Kennwort in einem für die Kennwortdatei geeigneten Format in die Standardausgabe. Wenn Sie beispielsweise den Benutzernamen *modeler* mit dem Kennwort "data mining" definieren wollen, geben Sie Folgendes ein:

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

Das Definieren eines einzigen Benutzernamens ist in den meisten Fällen, in denen sich alle Benutzer mit demselben Namen und Kennwort anmelden, ausreichend. Mit dem Operator *>>*, der an die einzelnen Dateien angehängt wird, können Sie jedoch noch weitere Benutzer erstellen, beispielsweise:

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

Hinweis: Mit einem einzigen *>* wird der Inhalt von *passwords.cfg* jedes Mal überschrieben und die zuvor festgelegten Benutzer werden ersetzt. Beachten Sie, dass unabhängig hiervon alle Benutzer dasselbe UNIX-Konto verwenden.

Hinweis: Wenn Sie der privaten Kennwortdatenbank neue Benutzer hinzufügen, während SPSS Modeler Server ausgeführt wird, müssen Sie SPSS Modeler Server neu starten, damit die neu definierten Benutzer erkannt werden. Ansonsten schlagen Anmeldeversuche aller neuen Benutzer, die seit dem Neustart von SPSS Modeler Server mit `pwutil` hinzugefügt wurden, fehl.

7. Ändern Sie die Besitzerrechte für das IBM SPSS Modeler-Installationsverzeichnis und des gesamten Inhalts rekursiv für den Benutzer `<modelerserv>` und die Gruppe `<modelerusers>`, wobei die zuvor von Ihnen definierten Namen referenziert werden. Beispiel:

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

8. Erstellen Sie Unterverzeichnisse im Datenverzeichnis für Ihre IBM SPSS Modeler-Benutzer, sodass sie einen Ort haben, an dem sie störungsfrei Ihre Arbeitsdaten speichern können. Diese Verzeichnisse sollten im Besitz der Gruppe `<modelerusers>` sein, die über Lese-, Schreib- und Suchberechtigungen für das Verzeichnis verfügen sollte. So erstellen Sie beispielsweise ein Arbeitsverzeichnis für den Benutzer `bob`

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

Außerdem können Sie für das Verzeichnis das Bit "set-group-ID" festlegen, sodass alle Datendateien, die in das Verzeichnis kopiert werden, automatisch in den Besitz der Gruppe `<modelerusers>` übergehen:

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

Ausführen von SPSS Modeler Server als Nicht-Root-Benutzer

So führen Sie SPSS Modeler Server als Nicht-Root-Benutzer aus:

1. Melden Sie sich mit dem zuvor erstellten Nicht-Root-Benutzerkonto an.
2. Wenn Sie bei der Ausführung die Konfigurationsdateioption `start_process_as_login_user` aktiviert haben, können Sie SPSS Modeler Server starten, stoppen und den Status überprüfen. Weitere Informationen finden Sie im Thema „Starten, Stoppen und Statusprüfung unter UNIX“ auf Seite 21.

Endbenutzer stellen eine Verbindung mit SPSS Modeler Server her, indem Sie sich über die Client-Software anmelden. Sie müssen Endbenutzern die nötigen Informationen zum Herstellen einer Verbindung übermitteln, inklusive der IP-Adresse oder dem Hostnamen des Server-Computers.

Ausführen von IBM SPSS Modeler Server als Nicht-Root-Benutzer unter AIX

Beim Start von IBM SPSS Modeler Server unter AIX als Non-Root-Benutzer besteht die Möglichkeit, dass der Fehler `lsattr not found` an der Konsole ausgegeben wird. Das liegt daran, dass `lsattr` sich im Verzeichnis `/usr/sbin` befindet, aber wenn IBM SPSS Modeler Server einen neuen Benutzer (als Non-Root) anmeldet, wird der Pfad so zurückgesetzt, dass er nur `/usr/bin` und `/bin` enthält.

Wenn dieser Fehler auftritt, legen Sie die folgende Verknüpfung unter `/usr/bin` an:

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

Starten Sie IBM SPSS Modeler Server nach dem Anlegen der Verknüpfung neu.

Fehlerbehebung bei fehlgeschlagener Benutzerauthentifizierung

Je nachdem, wie das Betriebssystem für die Durchführung der Authentifizierung konfiguriert ist, kann es bei der Ausführung in einer Nicht-Root-Konfiguration Probleme mit der Anmeldung bei SPSS Modeler Server geben. Dies kann beispielsweise dann vorkommen, wenn das Betriebssystem (mit der Datei `/etc/nsswitch.conf` o. ä.) so konfiguriert ist, dass die lokale Kennwortspiegeldatei überprüft wird, anstatt NIS oder LDAP zu verwenden.

Dies geschieht, da SPSS Modeler Server über Lesezugriff auf die Dateien verfügen muss, die zur Durchführung der Authentifizierung verwendet werden, einschließlich der Datei */etc/shadow* bzw. ihrer Entsprechung (z. B. */etc/security/passwd* unter AIX), in der sichere Informationen zum Benutzerkonto gespeichert werden. Die Dateiberechtigungen des Betriebssystems sind jedoch im Allgemeinen so festgelegt, dass nur der Root-Benutzer auf die Datei */etc/shadow* zugreifen kann. Unter diesen Umständen kann ein Nicht-Root-Prozess */etc/shadow* nicht zur Überprüfung der Benutzerkennwörter lesen, was zu einem Authentifizierungsfehler führt.

Es gibt mehrere Möglichkeiten zur Lösung dieses Problems:

- Bitten Sie den Systemadministrator, das Betriebssystem so zu konfigurieren, dass NIS oder LDAP für die Authentifizierung verwendet wird.
- Ändern Sie die Dateiberechtigungen für die geschützten Dateien, beispielsweise indem Sie Lesezugriff auf die Datei */etc/shadow* gewähren, sodass das lokale Benutzerkonto, das zur Ausführung von SPSS Modeler Server verwendet wird, auf die Datei zugreifen kann. Diese Problemumgehung mag in Produktionsumgebungen unpraktisch sein, sie könnte jedoch vorübergehend in einer Testumgebung angewendet werden, um zu überprüfen, ob das Fehlschlagen der Autorisierung mit der Betriebssystemkonfiguration zu tun hat.
- Geben Sie eine Zugriffssteuerungsliste (ACL) für die Datei */etc/shadow* an (siehe Beispiel weiter unten in diesem Abschnitt).
- Führen Sie SPSS Modeler Server als Root aus, damit die Serverprozesse die Datei */etc/shadow* lesen können. *Vorsicht:* Stellen Sie in diesem Fall sicher, dass die Datei *options.cfg* für SPSS Modeler Server die Option *start_process_as_login_user*, N beinhaltet, um die weiter oben erläuterten Sicherheitsprobleme zu vermeiden.

Beispiel für eine Zugriffssteuerungsliste (ACL)

Unter Solaris-Betriebssystemen können Sie mit dem Befehl *setfacl* eine Zugriffssteuerungsliste erstellen (halten Sie Rücksprache mit dem Administrator, um sicherzustellen, dass der Befehl auf Ihrem System verfügbar ist). Im folgenden Beispiel wird illustriert, wie eine solche Zugriffssteuerungsliste (ACL) für den Benutzer *modserv* erstellt werden kann:

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

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

# getfacl /etc/shadow

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

Anhang D. Konfigurieren und Ausführen von SPSS Modeler Server mit einer privaten Kennwortdatei unter Windows

Einführung

In den folgenden Anweisungen finden Sie Informationen zur Ausführung von IBM SPSS Modeler Server mit einer privaten Kennwortdatei auf Windows-Systemen. Bei dieser Methode authentifiziert IBM SPSS Modeler die Benutzer anhand einer privaten Kennwortdatenbank, die sich von der Systemauthentifizierung unter Windows unterscheidet.

Konfigurieren mit einer privaten Kennwortdatenbank

Wenn Sie auswählen, dass die Benutzer mittels einer privaten Kennwortdatenbank authentifiziert werden sollen, werden alle Sitzungen mit demselben Benutzerkonto gestartet.

So konfigurieren Sie IBM SPSS Modeler Server hierzu:

1. Erstellen Sie das Benutzerkonto, auf dem IBM SPSS Modeler Server ausgeführt werden soll. Dieses Konto dient zur ausschließlichen Verwendung für den IBM SPSS Modeler Server-Dämonprozess. Für dieses Beispiel wird der Name *modelerserv* verwendet.
2. Bearbeiten Sie die Datei mit den Serveroptionen *config/options.cfg*, um die Option `start_process_as_login_user` auf `Y` zu setzen und die Authentifizierung mit der privaten Kennwortdatenbank einzustellen, indem Sie die folgende Zeile anhängen:
`authentication_methods, "pasw_modeler"`
3. Anschließend erstellen Sie eine private Kennwortdatenbank, die in der Datei *config/passwords.cfg* gespeichert wird. Die Kennwortdatei definiert die Kombinationen aus Benutzername/Kennwort, die für die Anmeldung bei IBM SPSS Modeler zulässig sind. *Hinweis:* Diese Daten sind nur für IBM SPSS Modeler bestimmt (privat) und stehen nicht in Zusammenhang mit den Benutzernamen und Kennwörtern, die für die Anmeldung an Windows verwendet werden. Sie können dieselben Benutzernamen verwenden, wenn dies praktikabel ist, die Verwendung derselben Kennwörter ist jedoch nicht zulässig.

Zum Erstellen der Kennwortdatei benötigen Sie das Dienstprogramm für Kennwörter (*pwutil*), das sich im Verzeichnis *bin* der IBM SPSS Modeler Server-Installation befindet. Die Synopse dieses Programms lautet:

```
pwutil [ Benutzername [ Kennwort ] ]
```

Das Programm nimmt einen Benutzernamen und ein Nur-Text-Kennwort und schreibt den Benutzernamen und das verschlüsselte Kennwort in einem für die Kennwortdatei geeigneten Format in die Standardausgabe. Wenn Sie beispielsweise den Benutzernamen *modeler* mit dem Kennwort "data mining" definieren wollen, geben Sie Folgendes ein:

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

Das Definieren eines einzigen Benutzernamens ist in den meisten Fällen, in denen sich alle Benutzer mit demselben Namen und Kennwort anmelden, ausreichend. Mit dem Operator `>>`, der an die einzelnen Dateien angehängt wird, können Sie jedoch noch weitere Benutzer erstellen, beispielsweise:

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

Hinweis: Mit einem einzigen `>` wird der Inhalt von *passwords.cfg* jedes Mal überschrieben und die zuvor festgelegten Benutzer werden ersetzt. Beachten Sie, dass unabhängig hiervon alle Benutzer dasselbe UNIX-Konto verwenden.

Hinweis: Wenn Sie der privaten Kennwortdatenbank neue Benutzer hinzufügen, während SPSS Modeler Server ausgeführt wird, müssen Sie SPSS Modeler Server neu starten, damit die neu definierten Benutzer erkannt werden. Ansonsten schlagen Anmeldeversuche aller neuen Benutzer, die seit dem Neustart von SPSS Modeler Server mit `pwutil` hinzugefügt wurden, fehl.

4. Geben Sie dem in Schritt 1 erstellten Benutzer uneingeschränkten Zugriff auf die Datei mit den Serveroptionen *config/options.cfg* und das Verzeichnis *%ALLUSERSPROFILE%\IBM\SPSS*.
5. Stoppen Sie in den Systemservices den Service *IBM SPSS Modeler Server* und geben Sie für die Anmeldung (*Log on*) statt dem lokalen Systemkonto (*Local System Account*) das in Schritt 1 erstellte Benutzerkonto an. Starten Sie den Service anschließend neu.

Anhang E. Lastenausgleich mit Server-Clustern

Mit IBM SPSS Collaboration and Deployment Services kann das Plug-in Coordinator of Processes verwendet werden, um Dienste im Netz zu verwalten. Der Coordinator of Processes stellt Serververwaltungsfunktionen bereit, mit denen die Kommunikation und Verarbeitung zwischen Client und Server optimiert werden können.

Die zu verwaltenden Dienste, beispielsweise IBM SPSS Statistics Server oder IBM SPSS Modeler Server registrieren sich beim Start beim Coordinator of Processes und senden in regelmäßigen Abständen aktualisierte Statusnachrichten. Die Dienste können außerdem alle notwendigen Konfigurationsdateien in IBM SPSS Collaboration and Deployment Services Repository speichern und bei der Initialisierung abrufen.

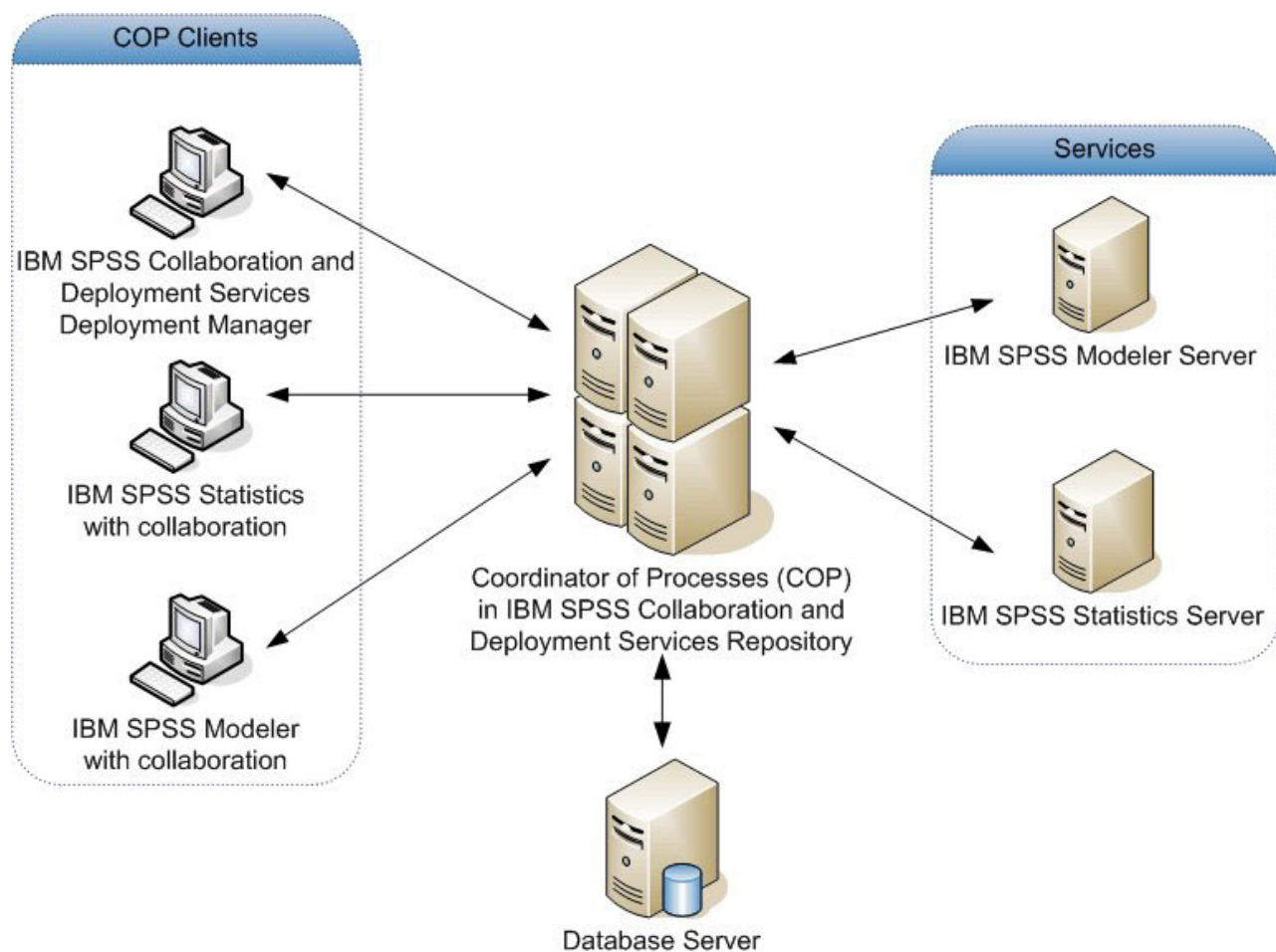


Abbildung 5. Architektur des Coordinator of Processes

Durch die Ausführung von IBM SPSS Modeler-Streams auf einem Server kann die Leistungsfähigkeit erhöht werden. In einigen Fällen haben Sie nur die Wahl zwischen einem Server und zwei Servern. In anderen Fällen kann eine größere Auswahl von Servern zur Verfügung stehen, da es einen substantziellen Unterschied zwischen den einzelnen Servern gibt, wie beispielsweise Inhaber, Zugriffsrechte, Serverdaten, ob es sich um einen Testserver oder einen Produktionsserver handelt usw. Wenn Coordinator of Processes in Ihrem Netz installiert ist, wird Ihnen möglicherweise zusätzlich ein Server-Cluster angeboten.

Ein Server-Cluster ist eine Gruppe von Servern, die hinsichtlich Konfiguration und Ressourcen austauschbar sind. Coordinator of Processes ermittelt, welcher Server am besten für die Antwort auf eine Verarbeitungsanforderung geeignet ist. Hierfür verwendet er einen Algorithmus, der die Last anhand mehrerer Kriterien ausgleicht, wie beispielsweise die Gewichtung der einzelnen Server, Benutzerprioritäten und die aktuelle Verarbeitungsbelastungen. Weitere Informationen finden Sie im Handbuch *Coordinator of Processes Service Developer Guide*, das auf Anfrage bei IBM zur Verfügung gestellt wird.

Wenn Sie in IBM SPSS Modeler eine Verbindung zu einem Server oder Server-Cluster herstellen, können Sie den Server manuell aufrufen oder mithilfe von Coordinator of Processes nach einem bestimmten Server oder Cluster suchen. Weitere Informationen finden Sie im Thema „Verbinden mit IBM SPSS Modeler Server“ auf Seite 15.

Bemerkungen

Diese Informationen wurden für weltweit angebotene Produkte und Dienstleistungen erarbeitet.

Möglicherweise bietet IBM die in dieser Dokumentation beschriebenen Produkte, Services oder Funktionen in anderen Ländern nicht an. Informationen über die gegenwärtig im jeweiligen Land verfügbaren Produkte und Services sind beim zuständigen IBM Ansprechpartner erhältlich. Hinweise auf IBM Lizenzprogramme oder andere IBM Produkte bedeuten nicht, dass nur Programme, Produkte oder Services von IBM verwendet werden können. Anstelle der IBM Produkte, Programme oder Services können auch andere, ihnen äquivalente Produkte, Programme oder Services verwendet werden, solange diese keine gewerblichen oder anderen Schutzrechte von IBM verletzen. Die Verantwortung für den Betrieb von Produkten, Programmen und Services anderer Anbieter liegt beim Kunden.

Für in diesem Handbuch beschriebene Erzeugnisse und Verfahren kann es IBM Patente oder Patentanmeldungen geben. Mit der Auslieferung dieses Handbuchs ist keine Lizenzierung dieser Patente verbunden. Lizenzanforderungen sind schriftlich an folgende Adresse zu richten (Anfragen an diese Adresse müssen auf Englisch formuliert werden):

IBM Director of Licensing
IBM Europe, Middle East & Africa
Tour Descartes
2, avenue Gambetta
92066 Paris La Defense
France

Trotz sorgfältiger Bearbeitung können technische Ungenauigkeiten oder Druckfehler in dieser Veröffentlichung nicht ausgeschlossen werden. Die hier enthaltenen Informationen werden in regelmäßigen Zeitabständen aktualisiert und als Neuausgabe veröffentlicht. IBM kann ohne weitere Mitteilung jederzeit Verbesserungen und/oder Änderungen an den in dieser Veröffentlichung beschriebenen Produkten und/oder Programmen vornehmen.

Verweise in diesen Informationen auf Websites anderer Anbieter werden lediglich als Service für den Kunden bereitgestellt und stellen keinerlei Billigung des Inhalts dieser Websites dar. Das über diese Websites verfügbare Material ist nicht Bestandteil des Materials für dieses IBM Produkt. Die Verwendung dieser Websites geschieht auf eigene Verantwortung.

Werden an IBM Informationen eingesandt, können diese beliebig verwendet werden, ohne dass eine Verpflichtung gegenüber dem Einsender entsteht.

Lizenznehmer des Programms, die Informationen zu diesem Produkt wünschen mit der Zielsetzung: (i) den Austausch von Informationen zwischen unabhängig voneinander erstellten Programmen und anderen Programmen (einschließlich des vorliegenden Programms) sowie (ii) die gemeinsame Nutzung der ausgetauschten Informationen zu ermöglichen, wenden sich an folgende Adresse:

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

Die Bereitstellung dieser Informationen kann unter Umständen von bestimmten Bedingungen - in einigen Fällen auch von der Zahlung einer Gebühr - abhängig sein.

Die Lieferung des in diesem Dokument beschriebenen Lizenzprogramms sowie des zugehörigen Lizenzmaterials erfolgt auf der Basis der IBM Rahmenvereinbarung bzw. der Allgemeinen Geschäftsbedingungen von IBM, der IBM Internationalen Nutzungsbedingungen für Programmpakete oder einer äquivalenten Vereinbarung.

Alle in diesem Dokument enthaltenen Leistungsdaten stammen aus einer kontrollierten Umgebung. Die Ergebnisse, die in anderen Betriebsumgebungen erzielt werden, können daher erheblich von den hier erzielten Ergebnissen abweichen. Einige Daten stammen möglicherweise von Systemen, deren Entwicklung noch nicht abgeschlossen ist. Eine Gewährleistung, dass diese Daten auch in allgemein verfügbaren Systemen erzielt werden, kann nicht gegeben werden. Darüber hinaus wurden einige Daten unter Umständen durch Extrapolation berechnet. Die tatsächlichen Ergebnisse können davon abweichen. Benutzer dieses Dokuments sollten die entsprechenden Daten in ihrer spezifischen Umgebung prüfen.

Alle Informationen zu Produkten anderer Anbieter stammen von den Anbietern der aufgeführten Produkte, deren veröffentlichten Ankündigungen oder anderen allgemein verfügbaren Quellen. IBM hat diese Produkte nicht getestet und kann daher keine Aussagen zu Leistung, Kompatibilität oder anderen Merkmalen machen. Fragen zu den Leistungsmerkmalen von Produkten anderer Anbieter sind an den jeweiligen Anbieter zu richten.

Aussagen über Pläne und Absichten von IBM unterliegen Änderungen oder können zurückgenommen werden und repräsentieren nur die Ziele von IBM.

Diese Veröffentlichung enthält Beispiele für Daten und Berichte des alltäglichen Geschäftsablaufs. Sie sollen nur die Funktionen des Lizenzprogramms illustrieren und können Namen von Personen, Firmen, Marken oder Produkten enthalten. Alle diese Namen sind frei erfunden; Ähnlichkeiten mit tatsächlichen Namen und Adressen sind rein zufällig.

Marken

IBM, das IBM Logo und ibm.com sind Marken oder eingetragene Marken der IBM Corp in den USA und/oder anderen Ländern. Weitere Produkt- und Servicenamen können Marken von IBM oder anderen Unternehmen sein. Eine aktuelle Liste der IBM Marken finden Sie auf der Webseite „Copyright and trademark information“ unter www.ibm.com/legal/copytrade.shtml.

Intel, das Intel-Logo, Intel Inside, das Intel Inside-Logo, Intel Centrino, das Intel Centrino-Logo, Celeron, Intel Xeon, Intel SpeedStep, Itanium und Pentium sind Marken oder eingetragene Marken der Intel Corporation oder der Tochtergesellschaften des Unternehmens in den USA und anderen Ländern.

Linux ist eine eingetragene Marke von Linus Torvalds in den USA, anderen Ländern oder beidem.

Microsoft, Windows, Windows NT und das Windows-Logo sind Marken der Microsoft Corporation in den USA und/oder anderen Ländern.

UNIX ist eine eingetragene Marke von The Open Group in den USA und anderen Ländern.

Java und alle auf Java basierenden Marken und Logos sind Marken oder eingetragene Marken der Oracle Corporation und/oder ihrer verbundenen Unternehmen.

Weitere Produkt- und Servicenamen können Marken von IBM oder anderen Unternehmen sein.

Index

Sonderzeichen

@SQLFN-Funktion 51

Numerische Stichwörter

64-Bit-Betriebssysteme 9

A

Administratorzugriff
für IBM SPSS Modeler Server 23
allow_modelling_memory_override
options.cfg, Datei 26
Anmeldung bei IBM SPSS Modeler Server 15
Anwendungsbeispiele 3
Arbeitsspeicher 10
Architektur
Komponenten 7
Authentifizierung 19
auto.sh (UNIX)
Speicherort von 55
Automatischer Serverstart
Konfiguration unter UNIX 55

B

Beispiele
Anwendungshandbuch 3
Übersicht 5
Benutzer-ID
IBM SPSS Modeler Server 15
Benutzerauthentifizierung 19
Benutzerkonten
IBM SPSS Modeler Server 19
Betriebssysteme
64-Bit 9

C

cache_compression
options.cfg, Datei 26
cache_connection, Option 30
Cachekomprimierung 26
Caching, datenbankintern 29
chemsrv.sh (UNIX)
Speicherort von 55
CLEM-Ausdrücke
SQL-Generierung 48
Coordinator of Processes 18
Lastenausgleich 69
Server-Cluster 69
Coordinator of Processes, Konfiguration
für IBM SPSS Modeler Server 28
COP 18
Lastenausgleich 69
Server-Cluster 69

cop_enabled
options.cfg, Datei 28
cop_host
options.cfg, Datei 28
COP-Konfiguration
für IBM SPSS Modeler Server 28
cop_password
options.cfg, Datei 28
cop_port_number
options.cfg, Datei 28
cop_service_description
options.cfg, Datei 28
cop_service_host
options.cfg, Datei 28
cop_service_name
options.cfg, Datei 28
cop_service_weight
options.cfg, Datei 28
cop_update_interval
options.cfg, Datei 28
cop_user_name
options.cfg, Datei 28

D

Data Access Pack
Fehlerbehebung für ODBC unter
UNIX 59
Konfigurieren von UNIX 56
ODBC, Konfiguration unter UNIX 56
und UNIX-Bibliothekspfade 61
data_file_path
options.cfg, Datei 25
data_files_restricted
options.cfg, Datei 25
Dateiberechtigungen
IBM SPSS Modeler Server 19
Konfiguration unter UNIX 56
Dateinamen
UNIX 12
Windows 12
Datenbankcaching
SQL-Generierung 43
Steuerung über options.cfg 29
Datenbanken
Zugriff 11
Datenbankinternes Caching 29
Datenbankinternes Mining 37
Datenbankserver 37
Datenbankverbindungen
schließen 30
Datendateien
IBM SPSS Statistics 12
Import und Export 12
Datenzugriff 11
DB2
SQL-Optimierung 39, 40
Dokumentation 3
Domänenname (Windows)
IBM SPSS Modeler Server 15

E

Ergebnisse
Datensatzreihenfolge 20
Runden von 20
Unterschiede zwischen Client und
Server 20

F

Fehler bei der Streamausführung 26
Festplatten 10
FIPS-Verschlüsselung 29
Firewall-Einstellungen
options.cfg, Datei 26

G

Geisterprozesse, IBM SPSS Modeler Server 22

H

Hardwareempfehlungen
für IBM SPSS Modeler Server 9
Hostname
IBM SPSS Modeler Server 15, 17

I

IBM SPSS Analytic Server
Konfigurationsoptionen 25
IBM SPSS Modeler 1
Dokumentation 3
IBM SPSS Modeler Administration Console 23
Administratorzugriff 23
IBM SPSS Modeler Server 1
Administratorzugriff 23
andere Ergebnisse als der Client 20
Benutzer-ID 15
Benutzerauthentifizierung 19
Benutzerkonten 19
Coordinator of Processes, Konfiguration 28
COP-Konfiguration 28
Dateierstellung 19
Domänenname (Windows) 15
Hostname 15, 17
Informationen für Endbenutzer 19
Kennwort 15
Konfigurationsoptionen 24
nicht reagierende Prozesse 22
Portnummer 15, 17, 25
Serverprozesse 29
Single Sign-on 16, 17
temporäres Verzeichnis 25
Überwachung 29
Verwaltung 23
Verwaltungsoptionen 22

- IBM SPSS Modeler Server starten
 - unter UNIX 21
 - unter Windows 21
- IBM SPSS Modeler Server stoppen
 - unter UNIX 21
 - unter Windows 21
- IBM SPSS Modeler Server-Verbindungen
 - hinzufügen 17, 18
- IBM SPSS Statistics-Datendateien
 - Import und Export 12
- IBM SPSS Statistics-Datenzugriffstechnologie 11
- io_buffer_size
 - options.cfg, Datei 26

K

- Kennwort
 - IBM SPSS Modeler Server 15
- Kernelgrenzwerte unter UNIX 21
- Knoten
 - Unterstützung von SQL-Generierung 44
- Knotencaching
 - in Datenbank schreiben 43
 - SQL-Generierung 43
- Konfigurationsoptionen
 - Anmeldeversuche 25
 - automatische SQL-Generierung 27
 - Coordinator of Processes 28
 - COP 28
 - Leistung und Optimierung 26
 - parallele Verarbeitung 26
 - Portnummer 25
 - Speicherverwaltung 26
 - SQL-Zeichenfolgenlänge 27
 - SSL-Datenverschlüsselung 28
 - Streamneuerstellung 26
 - temporäres Verzeichnis 25
 - Übersicht 24, 25
 - Verbindungen und Sitzungen 25
 - von IBM SPSS Modeler Server 22
 - Zugriff auf Datendateien 25

L

- Leistung
 - von IBM SPSS Modeler Server 35

M

- max_file_size
 - options.cfg, Datei 25
- max_login_attempts
 - options.cfg, Datei 25
- max_parallelism
 - options.cfg, Datei 26
- max_sessions
 - options.cfg, Datei 25
- max_sql_string_length
 - options.cfg, Datei 27
- Mehrere Streams, Ausführung 26
- memory_usage
 - options.cfg, Datei 26
- Microsoft SQL Server
 - SQL-Optimierung 39, 40

- Modellierung
 - Speicherverwaltung 26
- modelling_memory_limit_percentage
 - options.cfg, Datei 26
- Modellnuggets
 - SQL anzeigen für 43

N

- Nachrichten
 - generiertes SQL anzeigen 43

O

- ODBC
 - Konfiguration unter UNIX 56
- ODBC-Datenquellen
 - ODBC- und UNIX-Scripts 56
 - und UNIX 56
- Operatoren
 - SQL-Generierung 48
- Optimierung
 - SQL-Generierung 39, 40, 42
 - options.cfg, Datei 29
- Oracle
 - SQL-Optimierung 39, 40, 53

P

- Parallele Verarbeitung
 - Steuerung 26
- Pfade 12
- port_number
 - options.cfg, Datei 25
- Porteinstellungen
 - options.cfg, Datei 26
- Portnummer
 - IBM SPSS Modeler Server 15, 17, 25
- program_file_path
 - options.cfg, Datei 25
- program_files_restricted
 - options.cfg, Datei 25
- Protokolldateien
 - generiertes SQL anzeigen 43
- Prozesse, nicht reagierend 22
- Prozessoren 9
 - mehrfach 26
- Pushbacks 39, 40, 42, 44
 - CLEM-Ausdrücke 48

R

- RAM 10
- rc.modeler (UNIX)
 - Speicherort von 55

S

- Secure Sockets Layer 31
- Server
 - Anmeldung 15
 - nach COP für Server suchen 18
 - Single Sign-on 16, 17
 - Verbindungen hinzufügen 17

- Serverport, Einstellungen
 - options.cfg, Datei 26
- Sicherheit
 - Dateierstellung 19
 - Dateierstellung unter UNIX konfigurieren 56
 - SSL 31
- Single Sign-on 15
- Speicherplatz 10
 - berechnen 10
- Speicherverwaltung
 - Verwaltungsoptionen 26
- SQL
 - Abfragen 51
 - Anzeige für Modellnuggets 43
 - Doppelte Spaltennamen 51
 - Oracle optimieren 53
 - Vorschau für generiertes 43
- sql_generation_enabled
 - options.cfg, Datei 27
- SQL-Generierung 39, 40, 44
 - Aktivierung 42
 - Anzeige für Modellnuggets 43
 - CLEM-Ausdrücke 43, 48
 - für IBM SPSS Modeler Server aktivieren 27
 - Protokollierung 43
 - Streamneuerstellung 43
 - Tipps 43
 - Vorschau 43
- SQL-Pushback. Siehe auch "SQL-Generierung" 39
- SQL Server
 - SQL-Optimierung 39, 40
- SSL 31
 - Kommunikation schützen 31
 - Übersicht 31
- ssl_certificate_file
 - options.cfg, Datei 28
- SSL-Datenverschlüsselung
 - für IBM SPSS Modeler Server aktivieren 28
- ssl_enabled
 - options.cfg, Datei 28
- ssl_private_key_file
 - options.cfg, Datei 28
- ssl_private_key_password
 - options.cfg, Datei 28
- Status
 - von IBM SPSS Modeler Server unter UNIX 21
 - von IBM SPSS Modeler Server unter Windows 21
- stream_rewriting_enabled
 - options.cfg, Datei 26
- Streamneuerstellung 43
- Suche nach COP für Verbindungen 18

T

- temp_directory
 - options.cfg, Datei 25
- Temporäre Dateien 10
 - Berechtigungen für (IBM SPSS Modeler Server) 19
- Temporäres Verzeichnis
 - für IBM SPSS Modeler Server 25

U

UNC-Dateinamen 12

UNIX

Benutzerauthentifizierung 19

Bibliothekspfade 61

Dateiberechtigungen konfigurieren 56

Single Sign-on 17

UNIX-Kernelgrenzwerte 21

UNIX-Scripts

auto.sh 55

bearbeiten 56

modelersrv.sh 55

rc.modeler 55

V

Verbindungen

Server-Cluster 18

zu IBM SPSS Modeler Server 15, 17, 18

Verschlüsselung

FIPS 29

SSL 31

Verwaltung

von IBM SPSS Modeler Server 22

Violette Knoten

SQL-Optimierung 41

Vorschau

SQL-Generierung 43

