IBM SPSS Analytic Server Version3.5

Benutzerhandbuch

Hinweis

Vor Verwendung dieser Informationen und des darin beschriebenen Produkts sollten die Informationen unter „Bemerkungen” auf Seite 45 gelesen werden.

Produktinformation

Diese Ausgabe bezieht sich auf Version 3, Release 5, Modifikation 0 von IBM® SPSS Analytic Server und alle nachfolgenden Releases und Modifikationen, bis dieser Hinweis in einer Neuausgabe geändert wird.

Inhaltsverzeichnis

Kapitel 1. Analytic Server-Konsole

Analytic Server stellt eine Thin Client-Schnittstelle für die Verwaltung von Datenquellen und Projekten bereit.

Anmeldung

1. Geben Sie die URL von Analytic Server in die Adressleiste Ihres Browsers ein. Die URL können Sie von Ihrem Serveradministrator erhalten.
2. Geben Sie den Benutzernamen ein, mit dem die Anmeldung am Server erfolgen soll.
3. Geben Sie das Kennwort ein, das dem angegebenen Benutzernamen zugeordnet ist.

Anmerkung: Die Eingabe des Benutzernamens während der Anmeldeaufforderung der Analytic Server -Konsole erfolgt ohne das Suffix des Realmnamens. Infolgedessen wird Benutzern bei Verwendung mehrerer Realms die Dropdown-Liste **Realms** angezeigt, aus der diese den entsprechenden Realm auswählen können. Wenn nur ein Realm definiert ist, wird Benutzern die Dropdown-Liste **Realms** bei der Anmeldung bei Analytic Server nicht angezeigt.

Nach der Anmeldung wird die Startseite der Konsole angezeigt.

Navigieren in der Konsole

- In der Kopfzeile werden der Produktname und der Name des zurzeit angemeldeten Benutzers sowie der Link zum Hilfesystem angezeigt. Der Name des zurzeit angemeldeten Benutzers steht an oberster Stelle in einer Dropdown-Liste, die auch den Link für die Abmeldung enthält.
- Im Inhaltsbereich werden die Aktionen angezeigt, die Sie über die Startseite der Konsole ausführen können.

Datenquellen

Eine Datenquelle besteht aus einer Sammlung von Datensätzen und einem Datenmodell, die ein Dataset für die Analyse definieren. Die Quelle von Datensätzen kann eine Datei (Text mit Trennzeichen, Text mit fester Breite, Excel) in HDFS, eine relationale Datenbank, HCatalog oder georäumlich sein. Das Datenmodell definiert alle Metadaten (Feldnamen, Speicher, Messniveau usw.), die für die Analyse von Daten erforderlich sind. Datenquelleneigner können Zugriff auf Datenquellen erteilen oder einschränken.

Datenquellenliste

Die Hauptseite mit den Datenquellen enthält eine Liste mit Datenquellen, deren Mitglied der aktuelle Benutzer ist.

- Klicken Sie auf den Namen einer Datenquelle, um die zugehörigen Details anzuzeigen und die Eigenschaften zu bearbeiten.
- Geben Sie einen Suchbegriff in den Suchbereich ein, um die Liste zu filtern, damit nur Datenquellen angezeigt werden, deren Name den Suchbegriff enthält.
- Klicken Sie auf **Neu** , um eine neue Datenquelle mit dem Namen und Inhaltstyp zu erstellen, die Sie im Dialogfeld **Neue Datenquelle hinzufügen** angeben.
  - Informationen zu den Einschränkungen bei Namen, die Sie für Datenquellen vergeben können, finden Sie in „Benennungsregeln ” auf Seite 25.
  - Die verfügbaren Inhaltstypen sind File, Database, HCatalog und Geospatial.

    Hinweise:

    - Der Typ "HCatalog" ist nur verfügbar, wenn Analytic Server für das Arbeiten mit diesen Datenquellen konfiguriert wurde.

- Der Typ "HCatalog" ist für HDP 3.0 oder höher und für CDH 6.0 oder höher nicht verfügbar.
- Wenn der Inhaltstyp ausgewählt wurde, kann er nicht mehr bearbeitet werden.
- Sie können jetzt mehrere Datenquellen in einer einzelnen Aktion importieren/exportieren.

- Klicken Sie auf **Löschen** , um die Datenquelle zu entfernen. Bei dieser Aktion bleiben alle Dateien, die der Datenquelle zugeordnet sind, intakt.
- Klicken Sie auf **Aktualisieren** , um die Liste zu aktualisieren.
- Die ausgewählte Aktion wird in der Dropdown-Liste **Aktionen** ausgeführt.

1. Wählen Sie **Exportieren** aus, um ein Archiv aus den ausgewählten Datenquellen zu erstellen, und speichern Sie das Archiv im lokalen Dateisystem. Das Archiv enthält alle Dateien, die den ausgewählten Datenquellen im Modus **Projekte** oder **Datenquelle** hinzugefügt wurden.

   Anmerkung: Wenn nur eine Datenquelle ausgewählt ist, wird der Name der ausgewählten Datenquelle auch als Archivdateiname verwendet. Wenn mehrere Datenquellen ausgewählt sind, wird für die Archivdatei standardmäßig der Name `datasources.zip` angenommen.
2. Wählen Sie **Importieren** aus, um Archive zu importieren, die von der Exportaktion erstellt wurden.

   Anmerkung: Archivdateien, die Informationen aus mehreren Datenquellen enthalten, können nicht importiert werden. In diesen Fällen müssen zuerst die Archive der einzelnen Datenquellen aus dem Archiv `datasources.zip` extrahiert werden.
3. Wählen Sie **Duplizieren** aus, um eine Kopie der Datenquelle zu erstellen.

Individuelle Datenquellendetails

Der Inhaltsbereich ist in mehrere Abschnitte unterteilt, die vom Inhaltstyp der Datenquelle abhängen können.

Details zu
Diese Einstellungen sind für alle Inhaltstypen gleich.

Ihren Namen
Ein bearbeitbares Textfeld, in dem der Name der Datenquelle angezeigt wird.

Display name
Ein bearbeitbares Textfeld, in dem der Name der Datenquelle so wie in anderen Anwendungen angezeigt wird. Wenn dieses Feld leer ist, wird der in Name angegebene Name als Anzeigename verwendet.

Beschreibung
Ein bearbeitbares Textfeld, in dem Sie einen erläuternden Text zur Datenquelle angeben können.

Is public
Ein Kontrollkästchen, das angibt, ob jeder die Datenquelle sehen kann (Kästchen ist ausgewählt) oder ob Benutzer und Gruppen explizit als Mitglieder hinzugefügt werden müssen (Kästchen ist abgewählt).

Is global share
Ein Kontrollkästchen, das steuert, ob das Spark-RDD im globalen Cache gespeichert wird. Wenn diese Option ausgewählt ist, wird das Spark-RDD immer im globalen Cache gespeichert. Wenn diese Option inaktiviert ist, wird das Spark-RDD aus dem globalen Cache entfernt, wenn es von keinem Spark-Job verwendet wird.

Custom attributes
Anwendungen können durch Verwendung von angepassten Attributen Eigenschaften an Datenquellen anhängen und dadurch beispielsweise angeben, ob die Datenquelle temporär ist. Diese Attribute werden in der Analytic Server-Konsole verfügbar gemacht, um einen tieferen Einblick zu ermöglichen, wie Anwendungen die Datenquelle verwenden.

Klicken Sie auf **Speichern** , um den aktuellen Status der Einstellungen beizubehalten.

Sharing

Diese Einstellungen sind für alle Inhaltstypen gleich.

Sie können das Eigentumsrecht für eine Datenquelle freigeben, indem Sie Benutzer und Gruppen als Autoren oder Leser hinzufügen.

- Durch Eingeben eines Suchbegriffs in das Textfeld wird nach Benutzern und Gruppen gefiltert, deren Name den Suchbegriff enthält. Wählen Sie in der Dropdown-Liste **Autor** oder **Leser** aus, um ihre Rolle in der Datenquelle zuzuweisen. Klicken Sie auf **Mitglied hinzufügen** , um sie zur Liste der Mitglieder hinzuzufügen.
- Wählen Sie zum Entfernen eines Teilnehmers einen Benutzer oder eine Gruppe in der Mitgliederliste aus und klicken Sie auf **Mitglied entfernen**.

Anmerkung: Benutzer mit der Rolle **Administrator** haben Lese-und Schreibzugriff auf alle Datenquellen, unabhängig davon, ob sie speziell als Mitglied aufgelistet sind.

File Input

Einstellungen, die für die Definition von Datenquellen mit dem Inhaltstyp File spezifisch sind.

File Viewer

Hier werden für den Einschluss in die Datenquelle verfügbare Dateien angezeigt. Wählen Sie den Modus **Projects** aus, um Dateien in der Analytic Server-Projektstruktur anzuzeigen, wählen Sie **Data source** aus, um in einer Datenquelle gespeicherte Dateien anzuzeigen, oder wählen Sie **File system** aus, um das Dateisystem anzuzeigen (normalerweise HDFS). Sie können beide Ordnerstrukturen durchsuchen, aber HDFS kann überhaupt nicht bearbeitet werden. Im Modus **Projekte** können Sie keine Dateien hinzufügen, Ordner erstellen oder Elemente auf Stammebene löschen, sondern nur innerhalb definierter Projekte. Unter Projekte finden Sie Informationen zum Erstellen, Bearbeiten oder Löschen eines Projekts.

- Klicken Sie auf **Hochladen** , um eine Datei in die aktuelle Datenquelle bzw. das aktuelle Projekt bzw. den aktuellen Unterordner hochzuladen. Sie können in einem einzelnen Verzeichnis nach mehreren Dateien suchen und mehrere auswählen.

  Anmerkung: Dateien werden in das verteilte Dateisystem hochgeladen. Sie finden die hochgeladenen Dateien in der `/analytic-root`-Verzeichnisstruktur unter dem entsprechenden Nutzer, der entsprechenden Datenquelle oder dem entsprechenden Projekt (abhängig vom ausgewählten Modus) und dem entsprechenden Unterordner. Angenommen, Sie führen Folgendes aus:

  1. Anmelden am Nutzer `ibm`
  2. Erstellen einer Datenquelle `fraudDetection`
  3. Modus **Datenquelle** auswählen
  4. Erstellen eines Unterordners mit dem Namen `historicalData`
  5. Hochladen einer Datei `charges2015.csv`

  In diesem Fall befindet die Datei sich im verteilten Dateisystem in `/analytic-root/ibm/.datasource/fraudDetection/historicalData/charges2015.csv`. Angenommen jedoch, Sie führen Folgendes aus:

  1. Anmelden am Nutzer `ibm`
  2. Erstellen einer Datenquelle `fraudDetection`
  3. Auswählen des Modus **Projects**
  4. Auswählen eines vorhandenen Projekts `creditProcessing`
  5. Erstellen eines Unterordners mit dem Namen `historicalData`
  6. Hochladen einer Datei `charges2015.csv`

  In diesem Fall befindet die Datei sich im verteilten Dateisystem in `/analytic-root/ibm/creditProcessing/historicalData/charges2015.csv`.
- Klicken Sie auf **Neuer Ordner** , um einen neuen Ordner unter dem aktuellen Ordner mit dem Namen zu erstellen, den Sie im Dialogfenster "Neuer Ordnername" angeben.
- Klicken Sie auf **Download** , um die ausgewählten Dateien in das lokale Dateisystem herunterzuladen.

- Klicken Sie auf **Delete**, um die ausgewählten Dateien/Ordner zu entfernen.

Files included in data source definition
Verwenden Sie die Pfeilschaltfläche, um der Datenquelle ausgewählte Dateien oder Ordner hinzuzufügen oder daraus zu entfernen. Klicken Sie für jede ausgewählte Datei bzw. für jeden ausgewählten Ordner in der Datenquelle auf Settings, um die Spezifikationen für das Lesen der Datei zu definieren.

Wenn mehrere Dateien in einer Datenquelle vorhanden sind, müssen sie allgemeine Metadaten gemeinsam nutzen; das heißt, jede Datei muss dieselbe Anzahl Felder aufweisen, die Felder müssen in jeder Datei in derselben Reihenfolge geparst werden und jedes Feld muss über alle Dateien hinweg denselben Speicher belegen. Abweichungen zwischen Dateien können zur Folge haben, dass die Konsole die Vorschau und Metadaten Preview and Metadata nicht erstellen kann oder ansonsten gültige Werte als ungültig (null) geparst werden, wenn Analytic Server die Datei liest.

Database Selections

Geben Sie die Verbindungsparameter für die Datenbank an, die den Datensatzinhalt enthält.

Datenbank
Wählen Sie den Datenbanktyp aus, zu dem Sie eine Verbindung herstellen wollen. Folgendes steht zur Auswahl: Db2, Greenplum, Apache Impala, Amazon Redshift, MySQL, Netezza, Oracle, SQL Server, TeraData, Hive, DashDB oder BigSQL. Wenn der von Ihnen gesuchte Typ nicht aufgeführt ist, bitten Sie Ihren Serveradministrator, Analytic Server mit dem entsprechenden JDBC-Treiber zu konfigurieren.

Anmerkung: Analytic Server unterstützt MySQL-Datenbanken, die sich auf fernen Systemen befinden.

Hive Connect Type
Diese Option ist nur verfügbar, wenn `Hive` als Typ **Datenbank** ausgewählt ist. Wählen Sie den Verbindungstyp **Einzelserver** oder **Hochverfügbarkeit**aus. **Einzelserver** wird verwendet, wenn ein einzelner Hive -Server eingesetzt wird; **Hochverfügbarkeit** wird verwendet, wenn ein hoch verfügbarer Hive -Server-Cluster eingesetzt wird. Die folgenden Optionen sind verfügbar, wenn **Hochverfügbarkeit** ausgewählt ist:

ZooKeeper Quorum
Geben Sie eine durch Kommas getrennte Liste für alle ZooKeeper-Server im Format 'Host:Port' ein. Beispiel: `zkhost1:2181,zkhost2:2181`.

Name Space
Geben Sie den Hive-Stammnamensbereich in ZooKeeper ein. Beispiel: `hiveserver2` oder `hiveserver2-hive2` (wenn "hiveserver2 interactive" unter HDP 2.6 aktiviert ist und verwendet wird).

Hinweise:

- Die Werte für **Zookeeper Quorum** und **Namespace** befinden sich in der Datei `hive-site.xml` .
- Standardmäßig ist die Hive-Hochverfügbarkeit in Cloudera inaktiviert und muss manuell aktiviert werden.
- Wenn Sie eine Hive -Datenquelle in einer Umgebung verwenden, bei der es sich nicht um eineKerberos -Umgebung handelt, müssen Sie sicherstellen, dass die `Username` , die Sie im Abschnitt **Datenbankauswahl** eingegeben haben, dieselbe ist wie die Anmeldung AS `user`.

Server address
Geben Sie die URL des Servers an, auf dem sich die Datenbank befindet.

Server port
Die Nummer des Ports, an dem die Datenbank empfangsbereit ist.

Datenbankname
Der Name der Datenbank, zu der Sie eine Verbindung herstellen wollen.

Benutzername
Wenn die Datenbank kennwortgeschützt ist, geben Sie Ihren Benutzernamen ein.

Kennwort
Wenn die Datenbank kennwortgeschützt ist, geben Sie Ihr Kennwort ein.

Tabellenname
Geben Sie den Namen einer Tabelle aus der Datenbank ein, die Sie verwenden möchten.

Maximum concurrent reads
Geben Sie den Grenzwert für die Anzahl paralleler Abfragen ein, die von Analytic Server zur Datenbank gesendet werden können, um aus der in der Datenquelle angegebenen Tabelle zu lesen.

HCatalog Selections
Geben Sie die Parameter für den Zugriff auf Daten an, die unter Apache HCatalog verwaltet werden.

Datenbank
Der Name der HCatalog-Datenbank.

Tabellenname
Geben Sie den Namen einer Tabelle aus der Datenbank ein, die Sie verwenden möchten.

Filter
Der Partitionsfilter für die Tabelle, wenn die Tabelle als partitionierte Tabelle erstellt wurde. HCatalog-Filterung wird nur für Hive-Partitionsschlüssel mit dem Zeichenfolgetyp (string) unterstützt.

Anmerkung: Die Operatoren !=,<> und LIKE scheinen in bestimmten Hadoop-Verteilungen nicht zu funktionieren. Hierbei handelt es sich um ein Kompatibilitätsproblem zwischen HCatalog und den betreffenden Verteilungen.

HCatalog-Feldzuordnungen
Zeigt die Zuordnung eines Elements in HCatalog zu einem Feld in der Datenquelle an. Klicken Sie auf Edit, um die Feldzuordnungen zu ändern.

Anmerkung: Nach der Erstellung einer HCatalog-basierten Datenquelle, die Daten aus einer Hive-Tabelle bereitstellt, stellen Sie möglicherweise fest, dass Analytic Server das Lesen von Daten aus einer Datenquelle immer dann mit erheblicher Verzögerung beginnt, wenn die Hive-Tabelle aus einer großen Anzahl Dateien erstellt wird. Wenn Sie solche Verzögerungen feststellen, müssen Sie die Hive-Tabelle mit einer kleineren Anzahl von umfangreichen Datendateien erneut erstellen und die Anzahl Dateien dabei auf 400 oder weniger reduzieren.

Geospatial Selections
Geben Sie die Parameter für den Zugriff auf geografische Daten an.

Geospatial type
Die geografischen Daten können aus einem Online-Kartenservice oder einer Shapefile stammen.

Wenn Sie einen Kartenservice verwenden, geben Sie die URL des Service an und wählen Sie den gewünschten Kartenlayer aus.

Wenn Sie eine Shapefile verwenden, wählen Sie die Shapefile aus oder laden Sie sie hoch. Eine Shapefile ist ein Set von Dateien mit einem gemeinsamen Dateinamen, die in demselben Verzeichnis gespeichert werden. Wählen Sie die Datei mit dem Suffix SHP aus. Analytic Server sucht und verwendet die anderen Dateien. Es müssen immer zwei andere Dateien mit den Suffixen SHX und DBF vorhanden sein. Abhängig von der Shapefile können auch einige zusätzliche Dateien vorhanden sein.

Preview and Metadata
Nachdem Sie die Einstellungen für die Datenquelle angegeben haben, klicken Sie auf Preview and Metadata, um die Datenquellenspezifikationen zu prüfen und zu bestätigen..

Ausgabe
Datenquellen mit Datei-, Datenbank- oder HCatalog-Inhaltstyp können über die Ausgabe von Datenströmen angehängt oder überschrieben werden, die in Analytic Server ausgeführt werden. Wählen Sie **Schreibfähig machen** aus, um das Anhängen oder Überschreiben zu aktivieren und:

- Wählen Sie für Datenquellen mit Datenbankinhaltstyp eine Ausgabedatenbanktabelle aus, in die die Ausgabedaten geschrieben werden.
- Für Datenquellen mit Dateiinhaltstyp:
  1. Wählen Sie den Ausgabeordner aus, in den die neuen Dateien geschrieben werden.

     Tipp: Verwenden Sie für jede Datenquelle einen separaten Ordner, damit die Zuordnungen zwischen Dateien und Datenquellen leichter verfolgt werden können.
  2. Wählen Sie ein Dateiformat aus: **CSV** (durch Kommas getrennte Werte) oder **Splittable binary format**.
  3. Wählen Sie optional **Sequenzdatei erstellen**aus. Dies ist hilfreich, wenn Sie aufteilbare komprimierte Dateien erstellen wollen, die in nachfolgenden MapReduce-Jobs verwendet werden können.
  4. Wählen Sie **Zeilenumbrüche können mit Escapezeichen versehen werden** aus, wenn Ihre Ausgabe CSV ist und Sie Zeichenfolgefelder haben, die eingebettete Zeilenvorschub-oder Rücklaufzeichen enthalten. Dadurch wird jeder Zeilenumbruch als umgekehrter Schrägstrich gefolgt vom Buchstaben "n" geschrieben, ein Rücklauf wird als umgekehrter Schrägstrich gefolgt vom Buchstaben "r" und ein umgekehrter Schrägstrich wird als zwei aufeinanderfolgende umgekehrte Schrägstriche geschrieben. Solche Daten müssen mit derselben Einstellung gelesen werden. Es wird dringend empfohlen, bei der Verarbeitung von Zeichenfolgedaten, die Zeilenvorschub- oder Rücklaufzeichen enthalten, das Format Splittable binary format zu verwenden.
  5. Wählen Sie ein Komprimierungsformat aus. Die Liste enthält alle Formate, die zur Verwendung mit Ihrer Analytic Server-Installation konfiguriert wurden.

     Anmerkung: Manche Kombinationen aus Komprimierungsformat und Dateiformat führen dazu, dass die Ausgabe nicht aufgeteilt werden kann und die Ausgabe daher nicht für die weitere MapReduce-Verarbeitung geeignet ist. Analytic Server gibt eine Warnung im Abschnitt für die Ausgabe aus, wenn Sie eine solche Auswahl treffen.
- Wählen Sie für Datenquellen mit HCatalog-Inhaltstyp eine Hive-Ausgabetabelle aus, in die die Ausgabedaten geschrieben werden.

  Hinweise und Einschränkungen für HCatalog-Datenquellen:

  – Die HCatalog-Datenquellentabelle muss vorhanden sein, bevor mit Analytic Server gearbeitet wird (Analytic Server erstellt die erforderliche Tabelle nicht).
  – Das Metadaten-/Datenmodell der Tabelle muss mit dem Datenmodell der zu exportierenden Ergebnisse konsistent sein.
  – Die HCatalog-Datenquelle unterstützt nur den Anfügemodus; der Überschreibungsmodus wird nicht unterstützt.

Einstellungen (Dateidatenquellen)

Im Dialogfeld mit den Einstellungen (Settings) können Sie die Spezifikationen für das Lesen dateibasierter Daten definieren. Die Einstellungen gelten für alle ausgewählten Dateien und alle Dateien in den ausgewählten Ordnern, die den Kriterien auf der Registerkarte **Ordner** entsprechen.Die Angabe falscher Parsereinstellungen für eine Datei kann zur Folge haben, dass Vorschau und Metadaten von der Konsole nicht erstellt werden können oder eigentlich gültige Werte als ungültig (null) geparst werden, wenn Analytic Server die Datei liest.

Registerkarte "Settings"

Auf der Registerkarte Settings können Sie den Dateityp und die für den Dateityp spezifischen Parsereinstellungen angeben.

Sie können Datenquellen mithilfe von komprimierten Dateien für ein beliebiges unterstütztes Dateiformat definieren. Unterstützte Komprimierungsformate sind unter anderem Gzip, Deflate, Bz2, Snappy, and IBM CMX.

Typ für Dateien mit Trennzeichen

Dateien mit Trennzeichen sind Textdateien mit freien Feldern, deren Datensätze eine konstante Anzahl von Feldern, aber eine variable Anzahl von Zeichen pro Feld enthalten. Dateien mit Trennzeichen haben normalerweise die Dateierweiterung *.csv oder *.tab. Weitere Informationen finden Sie unter „Einstellungen für Dateitypen mit Trennzeichen" auf Seite 7.

Typ für Dateien mit festem Format

Textdateien mit festen Feldern sind Dateien, deren Felder nicht begrenzt sind, aber an derselben Position beginnen und eine feste Länge aufweisen. Textdateien mit festen Feldern haben normalerweise die Dateierweiterung *.dat. Weitere Informationen finden Sie unter „Einstellungen für unveränderliche Dateitypen" auf Seite 9.

Typ für semistrukturierte Dateien

Semistrukturierte Dateien (z. B. *.log) sind Textdateien, die eine vorhersehbare Struktur aufweisen, die über reguläre Ausdrücke Feldern zugeordnet werden kann. Diese Dateien sind jedoch nicht in dem hohen Maße strukturiert wie Dateien mit Trennzeichen. Weitere Informationen finden Sie unter „Einstellungen für semistrukturierte Dateitypen" auf Seite 10.

Text Analytics-Dateityp

Text Analytics-Dateien sind Dokumente (z. B. *.doc, *.pdf oder *.txt), die mit SPSS Text Analytics analysiert werden können.

Skip empty lines
Gibt an, ob leere Zeilen im extrahierten Textinhalt ignoriert werden sollen. Der Standardwert ist **Nein**.

Line separator
Gibt die Zeichenfolge an, mit der eine neue Zeile definiert wird. Standardwert ist das Zeilenvorschubzeichen "\n".

SPSS Statistics-Dateityp

SPSS Statistics-Dateien (*.sav, *.zsav) sind Binärdateien, die ein Datenmodell enthalten. Für diesen Dateityp sind keine weiteren Einstellungen auf der Registerkarte Settings erforderlich.

Typ für aufteilbare Binärformatdateien

Gibt an, dass es sich beim Dateityp um eine aufteilbare Datei im Binärformat (*.asbf) handelt. Dieser Dateityp kann alle Analytic Server-Feldtypen darstellen (im Unterschied zum Dateityp CSV, der Listenfelder nicht darstellen kann und spezielle Einstellungen für die Handhabung von integrierten Zeilenvorschub- und Rücklaufzeichen erfordert). Für diesen Dateityp sind keine weiteren Einstellungen auf der Registerkarte Settings erforderlich.

Typ für Sequenzdateien

Sequenzdateien (*.seq) sind Textdateien, die als Schlüssel/Wert-Paare strukturiert sind. Sie werden im Allgemeinen als intermediäres Format in MapReduce-Jobs verwendet.

Excel-Dateityp

Gibt an, dass es sich bei dem Dateityp um eine Microsoft Excel-Datei (*.xls, *.xlsx) handelt. Weitere Informationen finden Sie unter „Einstellungen für Excel-Dateitypen" auf Seite 11.

Einstellungen für Dateitypen mit Trennzeichen

Sie können die folgenden Einstellungen für Dateitypen mit Trennzeichen angeben.

Character set encoding
Die Zeichencodierung der Datei. Wählen Sie einen Java-Zeichensatznamen wie "UTF-8", "ISO-8859-2", "GB18030" usw. aus oder geben Sie diesen an. Der Standardwert ist **UTF-8**.

Field delimiters
Mindestens ein Zeichen, das Feldgrenzen markiert. Jedes Zeichen wird als unabhängiges Trennzeichen gesehen. Wenn Sie beispielsweise **Komma** und **Tabulator** auswählen (oder **Andere** auswählen und `,\t`eingeben), bedeutet dies, dass entweder ein Komma oder ein Tabulator die Feldbegrenzungen markiert. Wenn Steuerzeichen als Feldtrennzeichen fungieren, werden die hier angegebenen Zeichen zusätzlich zu den Steuerzeichen als Trennzeichen betrachtet. Wenn Steuerzeichen nicht als Feldtrennzeichen dienen, ist "," der Standardwert; andernfalls ist der Standardwert eine leere Zeichenfolge.

Control characters delimit fields
Legt fest, ob ASCII-Steuerzeichen, außer LF und CR, als Feldtrennzeichen betrachtet werden. Standardwert ist **No**.

First row contains field names
Legt fest, ob die erste Zeile für die Festlegung der Feldnamen verwendet werden soll. Standardwert ist **No**.

Number of initial characters to skip
Die Anzahl der Zeichen am Anfang der Datei, die übersprungen werden sollen. Eine nicht negative Ganzzahl. Der Standardwert ist "0" (null).

Merge white space
Legt fest, ob mehrere benachbarte Vorkommen eines Leerzeichens und/oder Tabulators als ein einziges Feldtrennzeichen betrachtet werden. Hat keine Auswirkung, wenn weder das Leerzeichen noch der Tabulator ein Feldtrennzeichen ist. Der Standardwert ist **Ja**.

End-of-line comment characters
Mindestens ein Zeichen, das Zeilenendekommentare markiert. Das Zeichen und alles, was im Datensatz darauf folgt, wird, ignoriert. Jedes Zeichen wird als unabhängige Kommentarmarkierung gesehen. "/*" bedeutet z. B., dass ein Kommentar entweder mit einem Schrägstrich oder einem Stern beginnt. Es ist nicht möglich, Kommentarmarkierungen aus mehreren Zeichen zu definieren, z. B. "//". Die leere Zeichenfolge signalisiert, dass keine Kommentarzeichen definiert sind. Wenn Kommentarzeichen definiert sind, werden diese überprüft, bevor Anführungszeichen verarbeitet oder zu überspringende Zeichen am Anfang übersprungen werden. Der Standardwert ist die leere Zeichenfolge.

Invalid characters
Legt fest, wie ungültige Zeichen (Bytesequenzen, die nicht Zeichen in der Codierung entsprechen) behandelt werden sollen.

Discard
Verwirft ungültige Bytesequenzen.

Replace with
Ersetzt jede ungültige Bytesequenz durch das angegebene einzelne Zeichen.

Single quotes
Gibt die Verarbeitung von einfachen Anführungszeichen (Hochkommas) an. Der Standardwert ist **Beibehalten**.

Keep
Hochkommas haben keine besondere Bedeutung und werden wie jedes andere Zeichen behandelt.

Drop
Hochkommas werden gelöscht, wenn sie nicht in Anführungszeichen stehen

Pair
Hochkommas werden als Anführungszeichen betrachtet und Zeichen zwischen zwei Hochkommas verlieren ihre besondere Bedeutung (sie werden als Zeichen in Anführungszeichen betrachtet). Ob einfache Anführungszeichen selbst in Zeichenfolgen in einfachen Anführungszeichen vorkommen können, wird durch die Einstellung **Anführungszeichen können durch Verdoppelung in Anführungszeichen gesetzt werden**bestimmt.

Double quotation marks
Gibt die Handhabung von Anführungszeichen an. Der Standardwert ist **Paar**.

Keep
Anführungszeichen haben keine besondere Bedeutung und werden wie jedes andere Zeichen behandelt.

Drop
Anführungszeichen werden gelöscht, wenn sie nicht in Anführungszeichen stehen

Pair
Anführungszeichen werden als Anführungszeichen betrachtet und Zeichen zwischen Paaren von Anführungszeichen verlieren ihre besondere Bedeutung (sie werden als Zeichen in Anführungszeichen betrachtet). Ob Anführungszeichen selbst innerhalb von Zeichenfolgen in Anführungszeichen auftreten können, wird durch die Einstellung **Anführungszeichen können durch Verdoppelung in Anführungszeichen gesetzt werden**bestimmt.

Quotes can be quoted by doubling
Gibt an, ob doppelte Anführungszeichen in Zeichenfolgen in doppelten Anführungszeichen und einfache Anführungszeichen in Zeichenfolgen in einfachen Anführungszeichen dargestellt werden können, wenn **Paar**angegeben ist. Bei Angabe von **Yes** werden Anführungszeichen innerhalb von Zeichenfolgen in Anführungszeichen und Hochkommas innerhalb von Zeichenfolgen in Hochkommas verdoppelt. Bei **Nein**gibt es keine Möglichkeit, ein Anführungszeichen innerhalb einer Zeichenfolge in doppelten Anführungszeichen oder ein einfaches Anführungszeichen innerhalb einer Zeichenfolge in einfachen Anführungszeichen zu setzen. Der Standardwert ist **Ja**.

Newlines can be escaped

Gibt an, ob der Parser einen umgekehrten Schrägstrich gefolgt vom Buchstaben "n" oder "r" oder einem weiteren umgekehrten Schrägstrich als Zeilenvorschubzeichen, Rücklaufzeichen oder als umgekehrten Schrägstrich interpretiert. Wenn Zeilenumbrüche nicht durch ein Escapezeichen entwertet sind, werden diese Zeichenfolgen einfach als umgekehrter Schrägstrich gefolgt vom Buchstaben "n" usw. gelesen. Der Standardwert ist **Nein**.

Einstellungen für unveränderliche Dateitypen
Sie können die folgenden Einstellungen für unveränderliche Dateitypen angeben:

Character set encoding
Die Zeichencodierung der Datei. Wählen Sie einen Java-Zeichensatznamen wie "UTF-8", "ISO-8859-2", "GB18030" usw. aus oder geben Sie diesen an. Der Standardwert ist **UTF-8**.

Invalid characters
Legt fest, wie ungültige Zeichen (Bytesequenzen, die nicht Zeichen in der Codierung entsprechen) behandelt werden sollen.

Discard
Verwirft ungültige Bytesequenzen.

Replace with
Ersetzt jede ungültige Bytesequenz durch das angegebene einzelne Zeichen.

Record length
Gibt an, wie Datensätze definiert werden. Bei **Zeilenumbruch mit Begrenzer**werden Datensätze durch Zeilenumbrüche, Dateianfang oder Dateiende definiert (begrenzt). Bei **Bestimmte Länge**werden Datensätze durch eine Satzlänge in Byte definiert. Geben Sie einen positiven Wert an.

Initial records to skip
Die Anzahl der Datensätze am Anfang der Datei, die übersprungen werden sollen. Geben Sie eine nicht negative ganze Zahl an. Der Standardwert ist 0.

Felder
In diesem Abschnitt werden die Felder in der Datei definiert. Klicken Sie auf **Feld hinzufügen** und geben Sie den Feldnamen, die Spalte, in der die Feldwerte beginnen, und die Länge der Feldwerte an. Spalten werden in einer Datei mit null beginnend nummeriert.

Einstellungen für semistrukturierte Dateitypen

Einstellungen für semistrukturierte Dateien bestehen aus Regeln für die Zuordnung des Dateiinhalts zu Feldern.

Rules Table

Einzelne Regeln extrahieren Informationen aus einem Datensatz, um ein Feld zu erstellen. Kombiniert in einer Regeltabelle definieren Regeln alle Felder, die aus jedem Datensatz in einer Datenquelle extrahiert werden können.

Die Regeln in der Tabelle werden der Reihe nach auf jeden Datensatz angewendet. Wenn alle Regeln in der Tabelle mit dem Datensatz übereinstimmen, werden keine anderen Regeltabellen für die Verarbeitung des Datensatzes benötigt und es wird der nächste Datensatz verarbeitet. Wenn eine Regel in der Tabelle nicht übereinstimmt, werden alle durch vorherige Regeln in der Tabelle extrahierten Feldwerte verworfen. Falls eine andere Regeltabelle vorhanden ist, werden die Regeln in der betreffenden Tabelle auf den Datensatz angewendet. Wenn keine Tabelle mit dem Datensatz übereinstimmt, wird die Regel für Abweichungen (Mismatch) angewendet.

Mismatch

Sie können Datensätze, die keiner der Regeltabellen entsprechen, **überspringen** oder den Wert aller Felder im Datensatz auf **Fehlend** (null) setzen.

Export Rules

Sie können die zurzeit sichtbare Regeltabelle zwecks Wiederverwendung speichern. Die exportierte Tabelle wird auf dem Server gespeichert.

Import Rules

Sie können eine gespeicherte Regeltabelle in die zurzeit sichtbare Regeltabelle importieren. Dadurch werden alle von Ihnen für die betreffende Tabelle definierten Regeln überschrieben. Die beste Vorgehensweise besteht darin, eine neue Tabelle zu erstellen und dann eine Regeltabelle zu importieren.

Regeleditor:

Mit dem Regeleditor können Sie eine Extraktionsregel für ein einzelnes Feld erstellen.

Anonymous capture group

Eine Felderfassungsregel beginnt mit der Extraktion von Daten aus einem Datensatz normalerweise an der Position, an der die vorherige Regel endete. Wenn zwischen zwei Feldern in einer semistrukturierten Datenquelle irrelevante Informationen vorhanden sind, kann es deshalb sinnvoll sein, eine anonyme Erfassungsgruppe zu definieren, die den Parser an der Stelle positioniert, an der das nächste Feld beginnt. Wenn Sie **Anonyme Erfassungsgruppe**auswählen, werden die Steuerelemente für die Benennung und Beschriftung der Erfassungsgruppe inaktiviert, aber der Rest des Dialogfelds funktioniert normal.

Field name

Geben Sie einen Namen für das Feld ein. Er wird zum Definieren der Datenquellenmetadaten verwendet. Feldnamen müssen innerhalb einer Regeltabelle eindeutig sein.

Rule name

Geben Sie optional eine Beschriftung für die Regel ein.

Beschreibung

Geben Sie optional eine längere Beschreibung für die Regel ein.

Defining a rule

Es gibt zwei Methoden zum Definieren von Regeln.

Use controls for extraction rules

Die Verwendung von Steuerelementen vereinfacht die Erstellung von Extraktionsregeln.

1. Geben Sie den Punkt an, an dem mit der Extraktion von Felddaten begonnen werden soll. **Aktuelle Position** beginnt an der Stelle, an der die vorherige Regel gestoppt wurde, und **Überspringen bis** beginnt am Anfang des Datensatzes und ignoriert alle Zeichen, bis die im Textfeld angegebene Position erreicht ist. Wählen Sie **Einschließen** aus, wenn die Felddaten das Zeichen an der Anfangsposition enthalten sollen.
2. Wählen Sie eine Felderfassungsgruppe aus der Dropdown-Liste **Erfassen** aus.

3. Wählen Sie optional den Punkt aus, an dem die Extraktion von Felddaten gestoppt werden soll. **Leerzeichen** wird gestoppt, wenn Leerzeichen (z. B. Leerzeichen oder Tabulatoren) gefunden werden, und **Bei Zeichen** wird an der angegebenen Zeichenfolge gestoppt. Wählen Sie **Einschließen** aus, wenn die Felddaten das Zeichen an der Stoppposition enthalten sollen.

Manually define regexp rules
Wählen Sie diese Option aus, wenn Sie die Syntax für reguläre Ausdrücke selber schreiben wollen. Geben Sie im Textfeld **Regulärer Ausdruck** einen regulären Ausdruck ein.

Add Field Capture Group
Ermöglicht es Ihnen, den regulären Ausdruck zur späteren Verwendung zu speichern. Die gespeicherte Erfassungsgruppe wird in der Dropdown-Liste **Erfassen** angezeigt.

Der Regeleditor zeigt eine Vorschau der von dieser Regel aus dem ersten Datensatz extrahierten Daten an, nachdem alle vorherigen Regeln in der Regeltabelle angewendet wurden.

Einstellungen für Excel-Dateitypen
Sie können die folgenden Einstellungen für Excel-Dateien angeben.

Worksheet selection
Wählt das Excel-Arbeitsblatt als zu verwendende Datenquelle aus. Geben Sie entweder einen numerischen Index (der Index des ersten Arbeitsblatts ist 0) oder den Namen des Arbeitsblatts an. Standardmäßig wird das erste Arbeitsblatt verwendet.

Data range selection for import.
Sie können den Datenimport mit der ersten nicht leeren Zeile oder mit einem expliziten Zellenbereich beginnen.

- **Bereich beginnt in erster nicht leere Zeile.** Sucht die erste nicht leere Zelle und verwendet sie als linke obere Ecke des Datenbereichs.
- Geben Sie alternativ einen expliziten Zellenbereich nach Zeile und Spalte an. Wenn Sie beispielsweise den Excel-Bereich A1:D5 angeben wollen, können Sie A1 in das erste Feld und D5 in das zweite Feld eingeben (oder alternativ R1C1 und R5C4). Alle Zeilen im angegebenen Bereich werden zurückgegeben, einschließlich leerer Zeilen.

First row contains field names
Gibt an, ob die erste Zeile des ausgewählten Zellenbereichs die Feldnamen enthält. Der Standardwert ist **No**.

Stop reading after encountering blank rows
Gibt an, ob das Lesen von Datensätzen gestoppt wird, nachdem mehr als eine leere Zeile festgestellt wurde, oder ob das Lesen aller Daten bis an das Ende des Arbeitsblatts fortgesetzt wird, einschließlich leerer Zeilen. Der Standardwert ist **No**.

Registerkarte "Formats"

Auf der Registerkarte Formats können Sie Formatinformationen für die geparsten Felder definieren.

Feldkonvertierungseinstellungen

Trim white space
Entfernt Leerzeichen am Anfang und/oder Ende der Zeichenfolgefelder. Der Standardwert ist **None**. Die folgenden Werte werden unterstützt:

Keine
Entfernt Leerzeichen nicht.

Links
Entfernt Leerzeichen am Anfang der Zeichenfolge.

Rechts
Entfernt Leerzeichen am Ende der Zeichenfolge.

Both
Entfernt Leerzeichen am Anfang und Ende der Zeichenfolge.

Ländereinstellung
Definiert eine Ländereinstellung. Standardmäßig die Ländereinstellung des Servers. Die Ländereinstellungszeichenfolge sollte wie folgt angegeben werden: <Sprache>[_Land[_Variante]], wobei Folgendes gilt:

Sprache
Ein gültiger, aus zwei Buchstaben bestehender Code in Kleinbuchstaben gemäß ISO-639-Definition.

Land
Ein gültiger, aus zwei Buchstaben bestehender Code in Großbuchstaben gemäß ISO-3166-Definition.

Variante
Ein für den Anbieter oder Browser spezifischer Code.

Decimal separator
Legt das als Dezimaltrennzeichen verwendete Zeichen fest. Standardmäßig wird die für die Ländereinstellung spezifische Einstellung verwendet.

Grouping symbols
Legt fest, ob das für die Ländereinstellung spezifische Zeichen, das als Tausendertrennzeichen verwendet wird, verwendet werden soll.

Default date format
Definiert ein Standarddatumsformat. Es werden alle durch die Unicode-LDML-Spezifikation (LDML - Locale Data Markup Language) definierten Formatmuster unterstützt.

Default time format
Definiert ein Standardzeitformat.

Default timestamp
Definiert ein Standardzeitmarkenformat.

Default time zone
Legt die Zeitzone fest. Standardmäßig wird UTC verwendet. Die Einstellung gilt für die Zeit- und Zeitmarkenfelder, für die nicht explizit eine Zeitzone angegeben ist.

Field Overrides

In diesem Abschnitt können Sie Formatierungsanweisungen für einzelne Felder zuweisen. Wählen Sie ein Feld im Datenmodell aus oder geben Sie einen Feldnamen ein und klicken Sie auf **Hinzufügen** , um es zur Liste der Felder mit einzelnen Anweisungen hinzuzufügen. Klicken Sie auf **Entfernen** , um sie aus der Liste zu entfernen. Für ein in der Liste ausgewähltes Feld können Sie die folgenden Feldeigenschaften festlegen.

Storage
Legt den Speicherort des Felds fest.

Decimal separator
Legt für Realspeicherfelder das Zeichen fest, das als Dezimalzeichen verwendet wird. Standardmäßig wird die für die Ländereinstellung spezifische Einstellung verwendet.

Grouping symbols
Legt für Ganzzahl- oder Realspeicherfelder fest, ob das für die Ländereinstellung spezifische Tausendertrennzeichen verwendet werden soll.

Registerkarte "Formats"
Legt das Format für Datums-, Zeit- oder Zeitmarkenspeicherfelder fest. Wählen Sie ein Format in der Dropdown-Liste aus.

Registerkarte "Field Order"

Für Dateitypen mit Trennzeichen oder Excel-Dateitypen können Sie auf der Registerkarte Field Order die geparste Reihenfolge von Feldern für die Datei definieren. Dies ist wichtig, wenn mehrere Dateien in einer Datenquelle vorhanden sind, da die tatsächliche Reihenfolge der Felder über die Dateien hinweg

abweichen kann, aber die geparste Reihenfolge der Felder identisch sein muss, um ein konsistentes Datenmodell zu erstellen.Für unveränderliche und semistrukturierte Dateitypen wird die Reihenfolge auf der Registerkarte Settings definiert.

Wenn eine einzelne Datei in der Datenquelle vorhanden ist oder alle Dateien dieselbe Feldereihenfolge haben, können Sie die Standardeinstellung **Feldereihenfolge stimmt mit Datenmodell überein**verwenden. Wenn mehrere Dateien in der Datenquelle vorhanden sind und die Reihenfolge der Felder in der Datei nicht übereinstimmt, definieren Sie eine **spezifische Feldreihenfolge** für das Parsing der Datei.

1. Sie können der geordneten Liste ein Feld hinzuzufügen, indem Sie den Feldnamen eingeben oder aus der vom Datenmodell bereitgestellten Liste auswählen. Sie können alle Felder im Datenmodell gleichzeitig hinzufügen, indem Sie auf **Alle hinzufügen**klicken. Feldnamen werden der geordneten Liste nur ein einziges Mal hinzugefügt.
2. Verwenden Sie die Pfeiltasten, um die Felder wie gewünscht zu sortieren.

Wenn **Bestimmte Feldreihenfolge** verwendet wird, sind alle Felder, die nicht zur Liste hinzugefügt werden, nicht Teil der Ergebnismenge für diese Datei. Wenn im Datenmodell Felder vorhanden sind, die nicht in diesem Dialogfeld aufgelistet sind, werden die Werte im Ergebnisset als null angegeben.

Registerkarte "Folder"

Bei der Angabe von Parsereinstellungen für einen Ordner können Sie auf der Registerkarte Folder auswählen, welche Dateien im Ordner in die Datenquelle eingeschlossen werden sollen.

Match all files in the selected folder
Die Datenquelle schließt alle in der höchsten Ebene des Ordners enthaltenen Dateien ein. Dateien in Unterordnern werden nicht eingeschlossen.

Match files using a regular expression
Die Datenquelle schließt alle in der höchsten Ebene des Ordners enthaltenen Dateien ein, die mit dem angegebenen regulären Ausdruck übereinstimmen.

Match files using a Unix globbing expression (potentially recursive)
Die Datenquelle schließt alle Dateien ein, die mit dem angegebenen UNIX-Globbing-Ausdruck übereinstimmen. Der Ausdruck kann Dateien enthalten, die sich in Unterordnern des ausgewählten Ordners befinden.

HCatalog-Feldzuordnungen

HCatalog Schema
Zeigt die Struktur der angegebenen Tabelle an. HCatalog kann ein hochgradig strukturiertes Dataset unterstützen. Wenn eine Analytic Server-Datenquelle für solche Daten definiert werden soll, muss die Struktur in einfache Zeilen und Spalten abgeflacht werden. Wählen Sie ein Element im Schema aus und klicken Sie auf die Pfeilschaltfläche, um es für die Analyse einem Feld zuzuordnen.

Nicht alle Baumknoten können zugeordnet werden. Beispielsweise wird ein Array oder eine Zuordnung komplexer Typen als "übergeordnetes Element" angesehen und kann nicht direkt zugeordnet werden; jedes einfache Element in einem HCatalog-Array oder in einer HCatalog-Zuordnung muss separat hinzugefügt werden. Diese Knoten sind durch eine Beschriftung im Baum gekennzeichnet, die auf "...:array:struct" oder auf "...:map:struct" endet.

Beispiel:

- Bei einem Array mit ganzen Zahlen können Sie ein Feld einem Wert innerhalb des Arrays zuweisen (bigintarray[45]), aber nicht das Array selbst (bigintarray).
- Bei einer Zuordnung können Sie ein Feld einem Wert innerhalb der Zuordnung zuweisen (datamap["Schlüssel"]), aber nicht die Zuordnung selbst (datamap).
- Bei einem Array mit einem Array mit ganzen Zahlen können Sie ein Feld einem Wert zuweisen (bigintarrayarray[45][2]), aber nicht das Array selbst (bigintarrayarray[45]).

  Wenn Sie ein Feld einem Array- oder Zuordnungselement zuweisen, muss die Definition des Elements deshalb den Index oder Schlüssel einschließen: bigintarray[Index] oder bigintmap["Schlüssel"].

Der aktuelle Benutzer kann nur die Tabellen sehen, auf die er/sie Zugriff hat. Da das HDFS-Verzeichnis das einzige Verzeichnis mit Lese- und Ausführungsberechtigung ist (die innere Datei hat Leseberechtigung, die der Benutzer sehen kann), können Benutzer Tabellen, auf die sie keinen Zugriff haben, nicht anzeigen. Diese Einschränkung dient zum Schutz von verwalteten Hive-Tabellen, externen Hive-Tabellen und partitionierten Verzeichnissen.

Feldzuordnungen

HCatalog Element
Doppelklicken Sie auf eine Zelle, um sie zu bearbeiten. Sie müssen die Zelle bearbeiten, wenn das HCatalog-Element ein Array oder eine Zuordnung ist. Geben Sie mit einem Array die ganze Zahl an, die dem Mitglied des Arrays entspricht, das Sie einem Feld zuordnen wollen. Geben Sie mit einer Zuordnung eine in Anführungszeichen eingeschlossene Zeichenfolge an, die dem Schlüssel entspricht, den Sie einem Feld zuordnen wollen.

Mapping Field
Das Feld, wie es in der Analytic Server-Datenquelle angezeigt wird. Doppelklicken Sie auf eine Zelle, um sie zu bearbeiten. Doppelte Werte in der Spalte Mapping Field sind nicht zulässig und führen zu einem Fehler.

Storage
Der Speicherort des Felds. Der Speicherort wird von HCatalog abgeleitet und kann nicht bearbeitet werden.

Anmerkung: Wenn Sie auf Preview and Metadata klicken, um eine HCatalog-Datenquelle fertigzustellen, stehen keine Bearbeitungsoptionen zur Verfügung.

Raw Data
Zeigt die Datensätze an, wie sie in HCatalog gespeichert sind. So können Sie leichter festlegen, wie das HCatalog-Schema Feldern zugeordnet werden soll.

Anmerkung: Jede bei HCatalog Selections angegebene Filterung wird auf die Ansicht mit den Rohdaten angewendet.

Verwenden von HCatalog-Datenquellen

Analytic Server bietet Unterstützung für HCatalog-Datenquellen. In diesem Abschnitt wird beschrieben, wie verschiedene zugrunde liegende NoSQL-Datenbanken eingerichtet werden.

In den meisten Fällen sollten Sie die Dokumentation des Anbieters zur Hive-Integration zu Rate ziehen.

Apache Accumulo
https://cwiki.apache.org/confluence/display/Hive/AccumuloIntegration

Apache Cassandra
„Apache Cassandra" auf Seite 14

Apache HBase
https://cwiki.apache.org/confluence/display/Hive/HBaseIntegration

MongoDB
https://github.com/mongodb/mongo-hadoop/wiki/Hive-Usage

Oracle NoSQL
https://docs.oracle.com/cd/E57371_01/doc.41/e57351/bigsql.htm#BIGUG21115

XML-Datenquellen
„XML-Datenquellen" auf Seite 16

Apache Cassandra

Analytic Server bietet Unterstützung für HCatalog-Datenquellen, denen Inhalt in Apache Cassandra zugrunde liegt.Cassandra stellt einen strukturierten Schlüssel/Wert-Speicher bereit. Schlüssel werden mehreren Werten zugeordnet, die als Spaltenfamilien gruppiert werden. Die Spaltenfamilien werden beim Erstellen einer Datenbank festgelegt, einer Familie können aber jederzeit Spalten hinzugefügt werden. Darüber hinaus werden Spalten nur angegebenen Schlüsseln hinzugefügt, sodass unterschiedliche

Schlüssel in jeder beliebigen Familie eine unterschiedliche Anzahl Spalten aufweisen können. Die Werte aus einer Spaltenfamilie für jeden Schlüssel werden zusammen gespeichert.

Cassandra-Tabellen können auf zwei Arten definiert werden: mit der traditionellen Cassandra-Befehlszeilenschnittstelle (cassandra-cli) und mit der neuen CQL-Shell (csqlsh).

Verwenden Sie die folgende Syntax, um eine externe Apache Cassandra-Tabelle in Hive zu erstellen, wenn die Tabelle mit der traditionellen Befehlszeilenschnittstelle erstellt wurde.

```
CREATE EXTERNAL TABLE <hive_table_name>  (<column specifications>)
STORED BY 'org.apache.hadoop.hive.cassandra.CassandraStorageHandler'
WITH SERDEPROPERTIES("cassandra.cf.name" = "<cassandra_column_family>",
"cassandra.host"="<cassandra_host>","cassandra.port" = "<cassandra_port>")
TBLPROPERTIES ("cassandra.ks.name" = "<cassandra_keyspace>");
```

Für die folgende CLI-Tabellendefinition beispielsweise

```
create keyspace test
with placement_strategy = 'org.apache.cassandra.locator.SimpleStrategy'
and strategy_options = [{replication_factor:1}];

create column family users with comparator = UTF8Type;

update column family users with
        column_metadata =
        [
        {column_name: first, validation_class: UTF8Type},
        {column_name: last, validation_class: UTF8Type},
        {column_name: age, validation_class: UTF8Type, index_type: KEYS}
        ];

assume users keys as utf8;

set users['jsmith']['first'] = 'John';
set users['jsmith']['last'] = 'Smith';
set users['jsmith']['age'] = '38';
set users['jdoe']['first'] = 'John';
set users['jdoe']['last'] = 'Dow';
set users['jdoe']['age'] = '42';

get users['jdoe'];
```

sieht die Hive-Tabellen-DDL wie folgt aus:

```
CREATE EXTERNAL TABLE cassandra_users  (key string, first string, last string, age string)
STORED BY 'org.apache.hadoop.hive.cassandra.CassandraStorageHandler'
WITH SERDEPROPERTIES("cassandra.cf.name" = "users",
"cassandra.host"="<cassandra_host>","cassandra.port" = "9160")
TBLPROPERTIES ("cassandra.ks.name" = "test");
```

Verwenden Sie die folgende Syntax, um eine externe Apache Cassandra-Tabelle in Hive zu erstellen, wenn die Tabelle mit CQL erstellt wurde.

```
CREATE EXTERNAL TABLE <hive_table_name>  (<column specifications>)
STORED BY 'org.apache.hadoop.hive.cassandra.cql.CassandraCqlStorageHandler'
WITH SERDEPROPERTIES("cassandra.cf.name" = "<cassandra_column_family>",
"cassandra.host"="<cassandra_host>","cassandra.port" = "<cassandra_port>")
TBLPROPERTIES ("cassandra.ks.name" = "<cassandra_keyspace>");
```

Für die folgende CQL3-Tabellendefinition beispielsweise

```
CREATE KEYSPACE TEST WITH REPLICATION = { 'class' : 'SimpleStrategy', 'replication_factor' :
2 };
USE TEST;

CREATE TABLE bankloan_10(
  row int,
  age int,
  ed int,
  employ int,
  address int,
  income int,
  debtinc double,
  creddebt double,
```

```
  othdebt double,
  default int,
  PRIMARY KEY(row)
);

INSERT INTO bankloan_10 (row, age,ed,employ,address,income,debtinc,creddebt,othdebt,default)
            VALUES (1,41,3,17,12,176,9.3,11.359392,5.008608,1);
INSERT INTO bankloan_10 (row, age,ed,employ,address,income,debtinc,creddebt,othdebt,default)
            VALUES (2,27,1,10,6,31,17.3,1.362202,4.000798,0);
INSERT INTO bankloan_10 (row, age,ed,employ,address,income,debtinc,creddebt,othdebt,default)
            VALUES (3,40,1,15,14,55,5.5,0.856075,2.168925,0);
INSERT INTO bankloan_10 (row, age,ed,employ,address,income,debtinc,creddebt,othdebt,default)
            VALUES (4,41,1,15,14,120,2.9,2.65872,0.82128,0);
INSERT INTO bankloan_10 (row, age,ed,employ,address,income,debtinc,creddebt,othdebt,default)
            VALUES (5,24,2,2,0,28,17.3,1.787436,3.056564,1);
INSERT INTO bankloan_10 (row, age,ed,employ,address,income,debtinc,creddebt,othdebt,default)
            VALUES (6,41,2,5,5,25,10.2,0.3927,2.1573,0);
INSERT INTO bankloan_10 (row, age,ed,employ,address,income,debtinc,creddebt,othdebt,default)
            VALUES (7,39,1,20,9,67,30.6,3.833874,16.668126,0);
INSERT INTO bankloan_10 (row, age,ed,employ,address,income,debtinc,creddebt,othdebt,default)
            VALUES (8,43,1,12,11,38,3.6,0.128592,1.239408,0);
INSERT INTO bankloan_10 (row, age,ed,employ,address,income,debtinc,creddebt,othdebt,default)
            VALUES (9,24,1,3,4,19,24.4,1.358348,3.277652,1);
INSERT INTO bankloan_10 (row, age,ed,employ,address,income,debtinc,creddebt,othdebt,default)
            VALUES (10,36,1,0,13,25,19.7,2.7777,2.1473,0);
```

sieht die Hive-Tabellen-DDL wie folgt aus:

```
CREATE EXTERNAL TABLE cassandra_bankloan_10 (row int, age int,ed int,employ int,address int,
            income int,debtinc double,creddebt double,othdebt double,default int)
STORED BY 'org.apache.hadoop.hive.cassandra.cql.CassandraCqlStorageHandler'
WITH SERDEPROPERTIES("cassandra.cf.name" = "bankloan_10","cassandra.host"="<cassandra_host>",
            "cassandra.port" = "9160")
TBLPROPERTIES ("cassandra.ks.name" = "test");
```

XML-Datenquellen

Analytic Server bietet Unterstützung für XML-Daten über HCatalog.

Beispiel

1. Ordnen Sie das XML-Schema den Hive-Datentypen über die Hive Data Definition Language (DDL) gemäß den folgenden Regeln zu:

   ```
   CREATE [EXTERNAL] TABLE <table_name> (<column_specifications>)
   ROW FORMAT SERDE "com.ibm.spss.hive.serde2.xml.XmlSerDe"
   WITH SERDEPROPERTIES (
       ["xml.processor.class"="<xml_processor_class_name>",]
       "column.xpath.<column_name>"="<xpath_query>",
        ...
       ["xml.map.specification.<element_name>"="<map_specification>"
        ...
       ]
   )
   STORED AS
       INPUTFORMAT "com.ibm.spss.hive.serde2.xml.XmlInputFormat"
       OUTPUTFORMAT "org.apache.hadoop.hive.ql.io.IgnoreKeyTextOutputFormat"
   [LOCATION "<data_location>"]
   TBLPROPERTIES (
       "xmlinput.start"="<start_tag ",
       "xmlinput.end"="<end_tag>"
   );
   ```

   Anmerkung: Wenn Ihre XML-Dateien mit Bz2-Komprimierung komprimiert werden, sollte `com.ibm.spss.hive.serde2.xml.SplittableXmlInputFormat` für `INPUTFORMAT` festgelegt werden. Wenn sie mit CMX-Komprimierung komprimiert werden, sollte `com.ibm.spss.hive.serde2.xml.CmxXmlInputFormat` festgelegt werden.

   Der folgende XML-Code beispielsweise...

   ```
   <records>
     <record customer_id="0000-JTALA">
       <demographics>
         <gender>F</gender>
         <agecat>1</agecat>
         <edcat>1</edcat>
         <jobcat>2</jobcat>
         <empcat>2</empcat>
         <retire>0</retire>
         <jobsat>1</jobsat>
         <marital>1</marital>
   ```

```
      <spousedcat>1</spousedcat>
      <residecat>4</residecat>
      <homeown>0</homeown>
      <hometype>2</hometype>
      <addresscat>2</addresscat>
    </demographics>
    <financial>
      <income>18</income>
      <creddebt>1.003392</creddebt>
      <othdebt>2.740608</othdebt>
      <default>0</default>
    </financial>
  </record>
</records>
```

würde durch die folgende Hive-DLL dargestellt werden:

```
CREATE TABLE xml_bank(customer_id STRING, demographics map<string,string>, financial map<string,string>)
ROW FORMAT SERDE 'com.ibm.spss.hive.serde2.xml.XmlSerDe'
WITH SERDEPROPERTIES (
    "column.xpath.customer_id"="/record/@customer_id",
    "column.xpath.demographics"="/record/demographics/*",
    "column.xpath.financial"="/record/financial/*"
)
STORED AS
    INPUTFORMAT 'com.ibm.spss.hive.serde2.xml.XmlInputFormat'
    OUTPUTFORMAT 'org.apache.hadoop.hive.ql.io.IgnoreKeyTextOutputFormat'
TBLPROPERTIES (
    "xmlinput.start"="<record customer",
    "xmlinput.end"="</record>"
);
```

Weitere Informationen finden Sie unter „Zuordnung von XML zu Hive-Datentypen” auf Seite 17.

2. Erstellen Sie eine Analytic Server-Datenquelle mit HCatalog-Inhaltstyp in der Analytic Server-Konsole.

Einschränkungen

- Zurzeit wird nur die XPath 1.0-Spezifikation unterstützt.
- Bei der Handhabung von Hive-Feldnamen wird nur der lokale Teil der qualifizierten Namen für die Elemente und Attribute verwendet. Die Namensbereichspräfixe werden ignoriert.

Zuordnung von XML zu Hive-Datentypen

Die in XML modellierten Daten können anhand der nachfolgend aufgeführten Konventionen in Hive-Datentypen transformiert werden.

Strukturen

Das XML-Element kann dem Hive-Strukturtyp direkt zugeordnet werden, sodass alle Attribute zu Dateneinträgen werden. Der Inhalt des Elements wird zu einem zusätzlichen Eintrag mit primitivem oder komplexem Typ.

XML-Daten

```
<result name="ID_DATUM">03.06.2009</result>
```

Hive-DDL und Rohdaten

```
struct<name:string,result:string>
```

```
{"name":"ID_DATUM", "result":"0.3.06.2009"}
```

Arrays

Die XML-Sequenzen von Elementen können als Hive-Arrays mit primitivem oder komplexem Typ dargestellt werden. Das folgende Beispiel zeigt, wie der Benutzer ein Array mit Zeichenfolgen unter Verwendung des Inhalts des XML-Elements <result> definieren kann.

XML-Daten

```
<result>03.06.2009</result>
<result>03.06.2010</result>
<result>03.06.2011</result>
```

Hive-DDL und Rohdaten

```
result array<string>
```

```
{"result":["03.06.2009","03.06.2010",...]}
```

Zuordnungen

Das XML-Schema stellt keine native Unterstützung für Zuordnungen bereit. Es gibt drei allgemeine Ansätze für die Modellierung von Zuordnungen in XML. Um den drei unterschiedlichen Ansätzen Rechnung zu tragen, wird die folgende Syntax verwendet:

```
"xml.map.specification.<element_name>"="<key>-><value>"
```

Wo

Elementname
Name des XML-Elements, das als Zuordnungseintrag berücksichtigt werden soll

Schlüssel
XML-Knoten für den Zuordnungseintragsschlüssel

Wert
XML-Knoten für den Zuordnungseintragswert

Die Zuordnungsspezifikation für das angegebene XML-Element sollte in der Hive-Tabellenerstellungs-DLL unter dem Abschnitt SERDEPROPERTIES definiert werden. Die Schlüssel und Werte können mithilfe der folgenden Syntax definiert werden:

@attribute
Mit der Spezifikation @attribute kann der Benutzer den Wert des Attributs als Schlüssel oder Wert für die Zuordnung verwenden.

Element ''
Der Elementname kann als Schlüssel oder Wert verwendet werden.

#content
Der Inhalt des Elements kann als Schlüssel oder Wert verwendet werden. Da die Zuordnungsschlüssel nur den primitiven Typ haben können, wird der komplexe Inhalt in eine Zeichenfolge konvertiert.

Die Ansätze zur Darstellung von Zuordnungen in XML und die entsprechende Hive-DLL sowie die entsprechenden Rohdaten werden nachfolgend beschrieben.

Elementname zu Inhalt
Der Name des Elements wird als Schlüssel und der Inhalt als Wert verwendet. Dies ist eines der gängigen Verfahren und wird standardmäßig beim Zuordnen von XML zu Hive-Zuordnungstypen verwendet. Die offensichtliche Einschränkung bei diesem Ansatz besteht darin, dass der Zuordnungsschlüssel nur den Zeichenfolgetyp haben kann.

XML-Daten

```
<entry1>value1</entry1>
<entry2>value2</entry2>
<entry3>value3</entry3>
```

Zuordnung, Hive-DDL und Rohdaten

In diesem Fall müssen Sie keine Zuordnung angeben, da standardmäßig der Name des Elements als Schlüssel und der Inhalt als Wert verwendet wird.

```
result map<string,string>
```

```
{"result":{"entry1": "value1", "entry2": "value2", "entry3": "value3"}}
```

Attribut zu Elementinhalt

Attributwert als Schlüssel und Elementinhalt als Wert verwenden.

XML-Daten

```
<entry name=”key1”>value1</entry>
<entry name=”key2”>value2</entry>
<entry name=”key3”>value3</entry>
```

Zuordnung, Hive-DDL und Rohdaten

```
"xml.map.specification.entry"="@name->#content"
```

```
result map<string,string>
```

```
{"result":{"key1": "value1", "key2": "value2", "key3": "value3"}}
```

Attribut zu Attribut

XML-Daten

```
<entry name=”key1” value=”value1”/>
<entry name=”key2” value=”value2”/>
<entry name=”key3” value=”value3”/>
```

Zuordnung, Hive-DDL und Rohdaten

```
"xml.map.specification.entry"="@name->@value"
```

```
result map<string,string>
```

```
{"result":{"key1": "value1", "key2": "value2", "key3": "value3"}}
```

Komplexer Inhalt

Komplexer Inhalt, der als Basiselementtyp verwendet wird, wird durch Hinzufügen eines Stammelements mit dem Namen < Zeichenfolge> in eine gültige XML-Zeichenfolge konvertiert. Beachten Sie die folgende XML:

```
<dataset>
 <value>10</value>
 <value>20</value>
 <value>30</value>
</dataset>
```

Der XPath-Ausdruck /dataset/* hat zur Folge, dass eine Reihe von XML-Knoten des Typs <value> zurückgegeben werden. Wenn das Zielfeld ein primitiver Typ ist, transformiert die Implementierung das Ergebnis der Abfrage in gültige XML, indem der Stammknoten <string> hinzugefügt wird.

```
<string>
 <value>10</value>
 <value>20</value>
 <value>30</value>
</string>
```

Anmerkung: Die Implementierung fügt kein Stammelement <string> hinzu, wenn das Ergebnis der Abfrage ein einzelnes XML-Element ist.

Textinhalt

Wenn ein XML-Element nur Leerzeichen als Text enthält, wird der Text ignoriert.

Vorschau und Metadaten (Datenquellen)

Wenn Sie auf **Vorschau und Metadaten** klicken, wird eine Stichprobe der Datensätze und des Datenmodells für die Datenquelle angezeigt. Hier können Sie die grundlegenden Metadateninformationen überprüfen.

Vorschau
Auf der Registerkarte Preview werden eine kleine Stichprobe der Datensätze und ihre Feldwerte angezeigt.

Bearbeiten

Auf der Registerkarte Edit werden die grundlegenden Feldmetadaten angezeigt. Für Datenquellen mit dem Inhaltstyp "Dateien" wird das Datenmodell anhand einer kleinen Stichprobe mit Datensätzen generiert. Sie können die Feldmetadaten auf dieser Registerkarte manuell bearbeiten. Für Datenquellen mit dem Inhaltstyp "HCatalog" wird das Datenmodell anhand der HCatalog-Feldzuordnungen generiert. Sie können den Feldspeicher auf dieser Registerkarte nicht bearbeiten.

Feld
Doppelklicken Sie auf den Feldnamen, um ihn zu bearbeiten.

Measurement
Dies ist das Messniveau, mit dem Merkmale der Daten in einem bestimmten Feld beschrieben werden.

Rolle
Wird verwendet, um Modellierungsknoten mitzuteilen, ob Felder für einen Computerlernprozess Eingabefelder (Input, Vorhersagefelder) oder Zielfelder (Target, vorherzusagende Felder) sind. Both und None sind ebenfalls verfügbare Rollen, ebenso wie die Rolle Partition, die ein Feld angibt, mit dem Datensätze zu Schulungs-, Test- und Prüfzwecken in separate Stichproben aufgeteilt werden. Der Wert Split gibt an, dass für jeden möglichen Wert des Feldes separate Modelle erstellt werden. Frequency gibt an, dass ein Feldwert als Häufigkeitsgewichtung für jeden Datensatz verwendet werden sollte. Mit Record ID wird ein Datensatz in der Ausgabe angegeben.

Storage
Mit der Option Storage wird beschrieben, wie Daten in einem Feld gespeichert werden. In einem Feld mit den Werten 1 und 0 werden z. B. ganzzahlige Daten gespeichert. Dies darf nicht mit dem Messniveau verwechselt werden, das die Verwendung der Daten beschreibt und sich nicht auf die Speicherung auswirkt. Sie können z. B. das Messniveau für ein Feld für Ganzzahlen mit den Werten 1 und 0 auf "Flag" setzen. Dies gibt normalerweise Folgendes an: 1 = True und 0 = False.

Werte
Zeigt die einzelnen Werte für Felder mit kategorialer Messung oder den Wertebereich für Felder mit fortlaufender Messung an.

Structure
Gibt an, ob Datensätze in dem Feld einen einzelnen Wert (primitives Element) oder eine Liste mit Werten enthalten.

Depth
Gibt die Tiefe einer Liste an; 0 ist eine Liste mit primitiven Elementen, 1 ist eine Liste mit Listen usw.

Scan all Data Values
Ermöglicht es Ihnen, einen Scan für die Datenwerte der Datenquelle einzuleiten und abzubrechen, um die Kategorienwerte und Bereichsgrenzwerte zu ermitteln. Wenn ein Scan in Bearbeitung ist, klicken Sie auf die Schaltfläche **Datenscan abbrechen**. Durch das Scannen aller Datenwerte wird

sichergestellt, dass die Metadaten korrekt sind. Allerdings kann das Scannen lange dauern, wenn die Datenquelle viele Felder und Datensätze enthält.

Projekte

Projekte sind Arbeitsbereiche, in denen Eingaben gespeichert werden und auf Ausgaben von Jobs zugegriffen wird. Sie stellen die Organisationsstruktur der höchsten Ebene bereit, die Dateien und Ordner enthält. Projekte können mit einzelnen Benutzern und Gruppen gemeinsam genutzt werden.

Projektliste

Die Hauptseite mit den Projekten enthält eine Liste mit Projekten, deren Mitglied der aktuelle Benutzer ist.

- Klicken Sie auf den Namen eines Projekts, um die zugehörigen Details anzuzeigen und die Eigenschaften zu bearbeiten.
- Geben Sie einen Suchbegriff in den Suchbereich ein, um die Liste zu filtern, damit nur Projekte angezeigt werden, deren Name den Suchbegriff enthält.
- Klicken Sie auf **Neu** , um ein neues Projekt mit dem im Dialog **Neues Projekt hinzufügen** angegebenen Namen zu erstellen. In „Benennungsregeln ” auf Seite 25 finden Sie Hinweise zu den Beschränkungen bei Namen, die Sie für Nutzer vergeben können.
- Klicken Sie auf **Löschen** , um die ausgewählten Projekte zu entfernen. Diese Aktion entfernt das Projekt und löscht alle zum Projekt gehörigen Daten aus HDFS.
- Klicken Sie auf **Aktualisieren** , um die Liste zu aktualisieren.

Einzelne Projektdetails

Der Inhaltsbereich ist in ausblendbare Abschnitte **Details**, **Teilen**, **Dateien**und **Versionen** unterteilt.

Details zu

Ihren Namen
Ein bearbeitbares Textfeld, in dem der Name des Projekts angezeigt wird.

Display name
Ein bearbeitbares Textfeld, in dem der Name des Projekts so wie in anderen Anwendungen angezeigt wird. Wenn dieses Feld leer ist, wird der in Name angegebene Name als Anzeigename verwendet.

Beschreibung
Ein bearbeitbares Textfeld, in dem Sie einen erläuternden Text zum Projekt angeben können.

Versions to keep
Löscht automatisch die älteste festgeschriebene Projektversion, wenn die Anzahl der Versionen die angegebene Anzahl überschreitet. Der Standardwert ist 25.

Anmerkung: Der Bereinigungsprozess wird nicht sofort, sondern alle 20 Minuten im Hintergrund ausgeführt.

Is public
Ein Kontrollkästchen, das angibt, ob jeder das Projekt sehen kann (Kästchen ist ausgewählt) oder ob Benutzer und Gruppen explizit als Mitglieder hinzugefügt werden müssen (Kästchen ist abgewählt).

Klicken Sie auf **Speichern** , um den aktuellen Status der Einstellungen beizubehalten.

Sharing
Sie können ein Projekt gemeinsam nutzen, indem Sie Benutzer und Gruppen als Autoren oder Anzeigeberechtigte hinzufügen.

- Durch Eingeben eines Suchbegriffs in das Textfeld wird nach Benutzern und Gruppen gefiltert, deren Name den Suchbegriff enthält. Wählen Sie die Ebene der gemeinsamen Nutzung aus und klicken Sie auf **Mitglied hinzufügen** , um zur Liste der Mitglieder hinzuzufügen.

  - Autoren sind Vollmitglieder eines Projekts und können das Projekt sowie die darin enthaltenen Ordner und Dateien ändern. Die Benutzer und Mitglieder dieser Gruppen haben Schreibzugriff (Analytic Server-Exportknoten) auf dieses Projekt, wenn sie über IBM SPSS Modeler eine Verbindung zu Analytic Server herstellen.
  - Anzeigeberechtigte können die Ordner und Dateien in einem Projekt anzeigen und Datenquellen über die Objekte in einem Projekt definieren, aber sie können das Projekt nicht ändern.
- Um einen Autor zu entfernen, wählen Sie einen Benutzer oder eine Gruppe in der Liste "Autor" aus und klicken Sie auf **Mitglied entfernen**.

Anmerkung: Administratoren verfügen über Lese- und Schreibzugriff auf jedes Projekt, unabhängig davon, ob sie namentlich als Mitglied aufgelistet sind.

Anmerkung: Auf der Registerkarte Sharing vorgenommene Änderungen werden sofort und automatisch angewendet.

Dateien

Projektstrukturbereich

Im rechten Bereich wird die Projekt-/Ordnerstruktur für das zurzeit ausgewählte Projekt angezeigt. Sie können die Ordnerstruktur durchsuchen, sie ist jedoch nur über die Schaltflächen bearbeitbar.

- Klicken Sie auf **Datei in das lokale Dateisystem herunterladen** , um eine ausgewählte Datei in das lokale Dateisystem herunterzuladen.
- Klicken Sie auf **Ausgewählte Datei (en) löschen** , um die ausgewählte Datei bzw. den ausgewählten Ordner zu entfernen.

File Viewer

Zeigt die Ordnerstruktur für das aktuelle Projekt an. Die Ordnerstruktur kann nur innerhalb von definierten Projekten bearbeitet werden. Das heißt, Sie können keine Dateien hinzufügen, Ordner erstellen oder Elemente auf der Stammebene des Modus **Projekte** löschen. Zum Erstellen oder Löschen müssen Sie zur Projektliste zurückkehren.

- Klicken Sie auf **Datei in HDFS** , um eine Datei in das aktuelle Projekt bzw. den aktuellen Unterordner hochzuladen.
- Klicken Sie auf **Neuen Ordner erstellen** , um unter dem aktuellen Ordner einen neuen Ordner mit dem im Dialog **Neuer Ordnername** angegebenen Namen zu erstellen.
- Klicken Sie auf **Datei in das lokale Dateisystem herunterladen** , um die ausgewählten Dateien in das lokale Dateisystem herunterzuladen.
- Klicken Sie auf **Ausgewählte Dateien löschen** , um die ausgewählten Dateien/Ordner zu entfernen.

Versions

Projekte werden auf der Basis von Änderungen des Datei- und Ordnerinhalts versioniert. Für Änderungen an den Attributen eines Projekts (z. B. die Beschreibung, ob es öffentlich ist und mit wem es gemeinsam genutzt wird) ist keine neue Version erforderlich. Zum Hinzufügen, Ändern oder Löschen von Dateien oder Ordnern ist eine neue Version erforderlich.

Tabelle für die Projektversionssteuerung

In der Tabelle werden die vorhandenen Projektversionen, deren Erstellungs- und Festschreibungsdatum, die Benutzer, die für die einzelnen Versionen verantwortlich sind, und die übergeordnete Version angezeigt. Die übergeordnete Version ist die Version, auf der die ausgewählte Version basiert.

- Klicken Sie auf **Sperren** , um Änderungen am Inhalt der ausgewählten Projektversion vorzunehmen.
- Klicken Sie auf **Commit**, um alle an einem Projekt vorgenommenen Änderungen zu speichern und diese Version zum aktuellen sichtbaren Status des Projekts zu machen.

- Klicken Sie auf **Verwerfen** , um alle Änderungen zu verwerfen, die an einem gesperrten Projekt vorgenommen wurden, und den sichtbaren Status des Projekts auf die zuletzt festgeschriebene Version zurückzugeben.
- Klicken Sie auf **Löschen** , um die ausgewählte Version zu entfernen.

Verwalten von Benutzern und Gruppen

Administratoren können über das Analytic Server-Benutzerverwaltungsportal Benutzer und Gruppen erstellen, löschen und ändern.

1. Erweitern Sie nach der Anmeldung bei der Analytic Server-Konsole das Dropdown-Menü neben Ihrer Anmelde-ID (in der Nähe der rechten oberen Ecke) und wählen Sie **User Management** aus. Das Analytic Server-Benutzerverwaltungsportal wird in einem separaten Browserfenster geöffnet.
2. Geben Sie Ihre Berechtigungsnachweise für die Analytic Server-Benutzerverwaltung ein und klicken Sie auf **Login**.

Analytic Server-Benutzerverwaltungsoptionen

Das Portal ist in die ausblendbaren Abschnitte **Manage User** und **Manage Group** unterteilt.

Manage User
Zu den Optionen gehören das Hinzufügen neuer Benutzer, das Bearbeiten vorhandener Benutzer, das Löschen vorhandener Benutzer und das Anzeigen der primären Benutzerliste.

Add User

Benutzername
Geben Sie einen gültigen Analytic Server-Benutzernamen ein.

Kennwort
Geben Sie in Kennwort für den angegebenen Benutzernamen ein.

Bestätigungskennwort
Geben Sie das Kennwort erneut ein.

Gruppen
Wählen Sie optional eine vorhandene Gruppe in der Dropdown-Liste aus. Nach dem Klicken auf **Abschicken**wird der Benutzer Mitglied der ausgewählten Gruppe.

User List
Zeigt alle Benutzer in einer Tabelle nach Benutzername und Gruppenzuordnung sortiert an.

- Verwenden Sie das Feld **Suchen** , um nach Benutzernamen in der Tabelle zu suchen.
- Verwenden Sie die Optionsfelder, um mehrere Benutzer auszuwählen. Klicken Sie auf **Löschen** , um die ausgewählten Benutzer zu löschen.
- Bearbeiten Sie Benutzereigenschaften, indem Sie auf einen Benutzernamen klicken.

  Anmerkung: Der Wert **Benutzername** kann nicht bearbeitet werden.

Manage Group
Zu den Optionen gehören das Hinzufügen neuer Gruppen, das Bearbeiten vorhandener Gruppen, das Löschen vorhandener Gruppen und das Anzeigen der primären Gruppenliste.

Add Group

Group Name
Geben Sie einen Gruppennamen ein.

Benutzer
Wählen Sie in der Dropdown-Liste Benutzer aus, die der Gruppe hinzugefügt werden sollen. Klicken Sie auf **Übergeben** , um die Gruppeneinstellungen zu speichern.

Group List
Zeigt alle Gruppen in einer Tabelle nach Gruppenname und Benutzern innerhalb jeder Gruppe sortiert an.

- Verwenden Sie das Feld **Suchen** , um nach Gruppennamen in der Tabelle zu suchen.
- Verwenden Sie die Optionsfelder, um mehrere Gruppen auszuwählen. Klicken Sie auf **Löschen** , um die ausgewählten Gruppen zu entfernen.
- Bearbeiten Sie Benutzereigenschaften, indem Sie auf einen Gruppennamen klicken.

  Anmerkung: Der Wert **Gruppenname** kann nicht bearbeitet werden.

Verwalten von Benutzer- und Gruppenrollen

Administratoren können die Rollen von Benutzern und Gruppen über die Seite Users verwalten.

Der Inhaltsbereich ist in ausblendbare Abschnitte **Details** und **Principals** unterteilt.

Details zu

Ihren Namen
Ein nicht bearbeitbares Textfeld, in dem der Name des Nutzers angezeigt wird.

Beschreibung
Ein bearbeitbares Textfeld, in dem Sie einen erläuternden Text zum Nutzer angeben können.

URL-Adresse
Die URL, mit der Benutzer sich über die Analytic Server-Konsole als Nutzer anmelden können.

Status
Aktive Tenants werden derzeit verwendet. Das Inaktivieren eines Tenants verhindert die Anmeldung von Benutzern bei diesem Tenant, löscht jedoch keine der zugrunde liegenden Informationen.

Principals

Principals sind Benutzer und Gruppen, die von dem Sicherheitsprovider übernommen werden, der während der Konfiguration konfiguriert wird. Sie können die Rolle von Principals in die Administrator-, Benutzer- oder Leserrolle ändern.

Metrics
Ermöglicht es Ihnen, Ressourcengrenzwerte für einen Nutzer zu konfigurieren. Gibt den zurzeit vom Nutzer belegten Plattenspeicherplatz zurück.

- Sie können eine Quote für den maximalen Plattenspeicherplatz für den Nutzer festlegen. Wenn dieser Grenzwert erreicht wird, können keine weiteren Daten für diesen Nutzer auf Platte geschrieben werden, bis genügend Plattenspeicherplatz freigegeben wird, damit die Plattenspeicherplatzbelegung des Nutzers unter die Quote fällt.
- Sie können eine Warnstufe für den Plattenspeicherplatz des Nutzers festlegen. Wenn die Quote überschritten wird, können von Principals keine Analysejobs für diesen Nutzer übergeben werden, bis genügend Plattenspeicherplatz freigegeben wird, damit die Plattenspeicherplatzbelegung des Nutzers unter die Quote fällt.
- Sie können eine maximale Anzahl paralleler Jobs festlegen, die gleichzeitig für diesen Nutzer ausgeführt werden können. Wenn die Quote überschritten wird, können von Principals keine Analysejobs für diesen Nutzer übergeben werden, bis ein zurzeit ausgeführter Job abgeschlossen ist.
- Sie können die maximale Anzahl Felder festlegen, die eine Datenquelle haben kann. Dieser Grenzwert wird bei jedem Erstellen oder Aktualisieren einer Datenquelle geprüft.
- Sie können die maximale Anzahl Datensätze festlegen, die eine Datenquelle haben kann. Dieser Grenzwert wird bei jedem Erstellen oder Aktualisieren einer Datenquelle geprüft; z. B. wenn Sie eine neue Datei hinzufügen oder Einstellungen für eine Datei ändern.
- Sie können die maximale Dateigröße in Megabyte festlegen. Dieser Grenzwert wird beim Hochladen einer Datei geprüft.

Security provider configuration
Hier können Sie den Provider für die Benutzerauthentifizierung angeben. **Standard** verwendet den Provider des Standardtenants, der während der Installation und Konfiguration eingerichtet wurde. Bei Angabe von **LDAP** können Sie Benutzer über einen externen LDAP-Server wie beispielsweise Active

Directory oder OpenLDAP authentifizieren. Geben Sie die Einstellungen für den Provider und optional Filtereinstellungen an, um die im Abschnitt Principals verfügbaren Benutzer und Gruppen zu steuern.

Benennungsregeln

Bei allen Elementen, für die ein eindeutiger Name in Analytic Server vergeben werden kann, z. B. Datenquellen und Projekte, gelten die folgenden Regeln für Namen:

- Innerhalb eines Nutzers müssen Namen in Objekten desselben Typs eindeutig sein. Beispielsweise kann nicht für zwei Datenquellen der Name insuranceClaims vergeben werden, aber eine Datenquelle und ein Projekt könnten jeweils den Namen insuranceClaims erhalten.
- Bei Namen muss Groß-/Kleinschreibung beachtet werden. insuranceClaims und InsuranceClaims beispielsweise werden als eindeutige Namen betrachtet.
- Bei Namen werden führende und abschließende Leerzeichen ignoriert.
- Die folgenden Zeichen sind in Namen ungültig:

```
~, #, %, &, *, {, }, \\, :, <, >, ?, /, |, ", \t, \r, \n
```

Kapitel 2. SPSS Modeler-Integration

SPSS Modeler ist eine Data-Mining-Workbench, die über einen visuellen Ansatz für die Analyse verfügt. Jede einzelne Aktion in einem Job, vom Zugriff auf eine Datenquelle über die Zusammenführung von Datensätzen bis zur Ausgabe einer neuen Datei oder zur Erstellung eines Modells, wird durch einen Knoten im Erstellungsbereich dargestellt. Diese Aktionen werden miteinander verknüpft, damit ein Analysedatenstrom gebildet wird. So erstellen Sie einen Analysedatenstrom, der mit Analytic Server ausgeführt wird:

1. Der Datenstrom muss mit einem Analytic Server-Quellenknoten beginnen.
2. Erstellen Sie den mittleren Teil des Datenstroms wie üblich in der Modeler-Schnittstelle, indem Sie von Analytic Server unterstützte Prozessknoten (Feld- oder Datensatzoperationen) auswählen. Die Modeller-Palette enthält eine Analytic Server-Anzeige mit den unterstützten Knoten.
3. Für die Fertigstellung des Datenstroms gibt es mehrere Optionen.
   - Wählen Sie einen von Analytic Server unterstützten Endknoten (Ausgabe, Diagramm, Export oder Modellierung) aus. In diesem Fall übertragt Modeler den gesamten Datenstrom mit einer Push-Operation an Analytic Server. Analytic Server koordiniert die erforderlichen Jobs im Hadoop-Cluster und stellt Modeler die Ergebnisse zur Verfügung. Modeler verwendet die Ergebnisse und stellt sie dar, als wäre der Datenstrom lokal verarbeitet worden.
   - Wenn Sie einen Endknoten auswählen, der von Analytic Server nicht unterstützt wird, überträgt Modeler so viel wie möglich vom Datenstrom mit einer Push-Operation an Analytic Server und beginnt dann mit dem Extrahieren von Datensätzen aus Hadoop (Pull-Operation). Beachten Sie, dass Analytic Server ein Scoring für Modelle durchführen kann, die zurzeit mit Analytic Server nicht erstellt werden können. Das heißt, dass Sie einen Datenstrom so strukturieren können, dass mit Analytic Server eine statistisch gültige Teilstichprobe Ihrer Big Data gezogen und anschließend mit Modeller "lokal" ein Modell erstellt wird. Das resultierende Modellnugget kann dann in einen Scoring-Datenstrom eingeschlossen werden, der vollständig in Analytic Server ausgeführt wird.

   Anmerkung: Sie können die maximale Anzahl der Datensätze festlegen, die SPSS Modeler in den Analytic Server-Datenstromeigenschaften aus Hadoop herunterlädt.

Unterstützte Knoten

Viele SPSS Modeler-Knoten werden für die Ausführung in HDFS unterstützt, bei der Ausführung bestimmter Knoten gibt es jedoch möglicherweise einige Unterschiede und einige Knoten werden zurzeit nicht unterstützt. In diesem Thema wird die aktuelle Unterstützungsstufe detailliert beschrieben.

Anmerkung: Informationen zur regulären Operation dieser Knoten finden Sie in der Dokumentation zu SPSS Modeler.

Allgemein

- Einige Zeichen, die normalerweise in einem Modeler-Feldnamen in Anführungszeichen zulässig sind, werden von Analytic Server nicht akzeptiert.
- Damit ein Modeler-Datenstrom in Analytic Server ausgeführt werden kann, muss er mit mindestens einem Analytic Server-Quellenknoten beginnen und mit einem einzelnen Modellierungsknoten oder Analytic Server-Exportknoten enden.
- Es wird empfohlen, den Speicher von stetigen Zielen als Speicher für reelle Zahlen und nicht als Speicher für ganze Zahlen festzulegen. Scoring-Modelle schreiben immer reelle Werte in die Ausgabedatendateien für stetige Ziele, während das Ausgabedatenmodell für die Scores dem Speicher des Ziels folgt. Wenn ein stetiges Ziel über einen Speicher für ganze Zahlen verfügt, gibt es daher eine Diskrepanz zwischen den geschriebenen Werten und dem Datenmodell für die Scores und diese Diskrepanz führt zu Fehlern, wenn Sie versuchen, die gescorten Daten zu lesen.

Quelle

- Ein Datenstrom, der mit etwas anderem als einem Analytic Server-Quellenknoten beginnt, wird lokal ausgeführt.

Datensatzoperationen

Alle Datensatzoperationen werden unterstützt, mit Ausnahme von Streaming-ZR- und Space-Time-Boxes-Knoten. Weitere Hinweise zur Funktionalität von unterstützten Knoten folgen.

Wählen

- Unterstützt dieselbe Funktionsgruppe wie der Ableitungsknoten.

Beispiel

- Stichprobenziehung auf Blockebene wird nicht unterstützt.
- Komplexe Methoden der Stichprobenziehung werden nicht unterstützt.
- Die erste n Stichprobenentnahme mit `"Discard sample"` wird nicht unterstützt.
- Die erste n Stichprobenentnahme mit `N>20000` wird nicht unterstützt.
- Stichprobenentnahme (1-in-n) wird nicht unterstützt, wenn `"Maximum sample size"` nicht festgelegt ist.
- Stichprobenentnahme (1-in-n) wird nicht unterstützt, wenn `N * "Maximum sample size" > 20000` angegeben ist.
- Stichprobenziehung "Zufällig %" auf Blockebene wird nicht unterstützt.
- "Zufällig %" unterstützt zurzeit die Angabe eines Startwerts.

Aggregieren

- Zusammenhängende Schlüssel werden nicht unterstützt. Wenn Sie einen vorhandenen Datenstrom wiederverwenden, der zum Sortieren von Daten konfiguriert ist, und diese Einstellung dann im Aggregatknoten verwenden, ändern Sie den Datenstrom, sodass der Sortierknoten entfernt wird.
- Reihenfolgestatistiken (Median, 1. Quartil, 3. Quartil) werden näherungsweise berechnet und über die Registerkarte für Optimierung unterstützt.

Sortieren

- Die Registerkarte für die Optimierung wird nicht unterstützt.

In einer verteilten Umgebung gibt es eine begrenzte Anzahl von Operationen, bei denen die vom Sortierknoten erstellte Datensatzreihenfolge beibehalten wird.

- Eine Sortierung, auf die ein Exportknoten folgt, erstellt eine sortierte Datenquelle.
- Ein Sortierknoten, auf den ein Stichprobenknoten mit der **ersten** Datensatzstichprobenziehung folgt, gibt die ersten *N* Datensätze zurück.

Im Allgemeinen sollten Sie einen Sortierknoten so nah wie möglich bei den Operationen platzieren, die die sortierten Datensätze benötigen.

Zusammenführen

- Das Zusammenführen nach Reihenfolge wird nicht unterstützt.
- Die Registerkarte für die Optimierung wird nicht unterstützt.
- Zusammenführungsoperationen sind relativ langsam. Wenn in HDFS Speicherplatz verfügbar ist, ist es unter Umständen weniger zeitintensiv, wenn Sie Ihre Datenquellen einmal zusammenführen und die zusammengeführte Quelle in den folgenden Datenströmen verwenden, anstatt die Datenquellen in jedem Datenstrom zusammenzuführen.

R-Transformation

Die R-Syntax im Knoten sollte aus Operationen bestehen, die jeweils nur für einen einzelnen Datensatz ausgeführt werden.

Feldoperationen

Alle Feldoperationen werden unterstützt, mit Ausnahme der Anonymisierungs-, Transponier-, Zeitintervall- und Verlaufsknoten. Weitere Hinweise zur Funktionalität von unterstützten Knoten folgen.

Autom. Datenvorbereitung

- Das Trainieren des Knotens wird nicht unterstützt. Die Anwendung der Transformationen in einem trainierten Knoten des Typs Autom. Datenvorbereitung auf neue Daten wird unterstützt.

Ableiten

- Alle Ableitungsfunktionen werden unterstützt, mit Ausnahme von Sequenzfunktionen.
- Das Ableiten eines neuen Felds als Anzahl ist im Grunde genommen eine Sequenzoperation und wird deshalb nicht unterstützt.
- Aufteilungsfelder können nicht in demselben Datenstrom abgeleitet werden, der sie als Aufteilungen verwendet. Sie müssen zwei Datenströme erstellen: einen, der das Aufteilungsfeld ableitet, und einen, der das Feld als Aufteilungen verwendet.

Füller

- Unterstützt dieselbe Funktionsgruppe wie der Ableitungsknoten.

Klassierung

Die folgende Funktionalität wird nicht unterstützt:

- Optimales Klassieren
- Ränge
- N-Perzentile -> Perzentilmethode: Summe der Werte
- N-Perzentile -> Perzentilmethode: "In aktuellem beibehalten" und "Zufällig zuweisen"
- N-Perzentile -> Benutzerdefiniert N: Werte über 100 und jeder N-Wert, bei dem 100 % N ungleich Null ist

RFM-Analyse

- Die Option "In aktuellem beibehalten" für die Handhabung von Bindungen wird nicht unterstützt. RFM-Scores (Recency, Frequency, Monetary - Aktualität, Häufigkeit, Geldwert) stimmen nicht immer mit denen überein, die von Modeler aus denselben Daten berechnet werden. Die Scorebereiche sind identisch, Scorezuweisungen (Klassennummern) können sich jedoch um 1 unterscheiden.

Diagramme

Alle Diagrammknoten werden unterstützt.

Modellierung

Die folgenden Modellierungsknoten werden unterstützt: Zeitreihen, TCM, Tree-AS, C&R-Baum, Quest, CHAID, Linear, Linear-AS, Neuronales Netz, GLE, LSVM, TwoStep-AS, Random Trees, STP und Assoziationsregeln. Weitere Hinweise zur Funktionalität dieser Knoten folgen.

Linear

Beim Erstellen von Modellen für große Datenmengen und -vielfalt wird das Ziel normalerweise in "Sehr große Datasets" geändert oder es werden Aufteilungen angegeben.

- Fortlaufendes Training vorhandener PSM-Modelle wird nicht unterstützt.
- Das Modellerstellungsziel Standard wird nur empfohlen, wenn Aufteilungsfelder so definiert sind, dass die Anzahl an Datensätzen in den einzelnen Aufteilungen nicht "zu groß" ist, wobei die Definition von "zu groß" von der Leistungsstärke einzelner Knoten in Ihrem Hadoop-Cluster abhängt. Im Gegensatz dazu müssen Sie auch darauf bedacht sein, sicherzustellen, dass Aufteilungen nicht so fein definiert sind, dass zu wenige Datensätze für die Erstellung eines Modells vorhanden sind.
- Das Ziel Boosting wird nicht unterstützt.
- Das Ziel Bagging wird nicht unterstützt.

- Das Ziel Sehr große Datasets wird nicht empfohlen, wenn wenige Datensätze vorhanden sind. Oft wird dann entweder kein Modell oder ein vermindertes Modell erstellt.
- Die automatische Datenaufbereitung wird nicht unterstützt. Dies kann Probleme verursachen, wenn versucht wird, anhand von Daten mit vielen fehlenden Werten ein Modell zu erstellen. Normalerweise würden diese als Teil der automatischen Datenaufbereitung imputiert. Als Problemumgehung kann ein Baummodell oder ein neuronales Netz mit der Einstellung Erweitert verwendet werden, um fehlende ausgewählte Werte zu imputieren.
- Die Genauigkeitsstatistik wird für aufgeteilte Modelle nicht berechnet.

Neuronales Netz

Beim Erstellen von Modellen für große Datenmengen und -vielfalt wird das Ziel normalerweise in "Sehr große Datasets" geändert oder es werden Aufteilungen angegeben.

- Fortlaufendes Training vorhandener Standard- oder PSM-Modelle wird nicht unterstützt.
- Das Modellerstellungsziel Standard wird nur empfohlen, wenn Aufteilungsfelder so definiert sind, dass die Anzahl an Datensätzen in den einzelnen Aufteilungen nicht "zu groß" ist, wobei die Definition von "zu groß" von der Leistungsstärke einzelner Knoten in Ihrem Hadoop-Cluster abhängt. Im Gegensatz dazu müssen Sie auch darauf bedacht sein, sicherzustellen, dass Aufteilungen nicht so fein definiert sind, dass zu wenige Datensätze für die Erstellung eines Modells vorhanden sind.
- Das Ziel Boosting wird nicht unterstützt.
- Das Ziel Bagging wird nicht unterstützt.
- Das Ziel Sehr große Datasets wird nicht empfohlen, wenn wenige Datensätze vorhanden sind. Oft wird dann entweder kein Modell oder ein vermindertes Modell erstellt.
- Wenn in den Daten viele Werte fehlen, verwenden Sie die Einstellung Erweitert, um fehlende Werte zu imputieren.
- Die Genauigkeitsstatistik wird für aufgeteilte Modelle nicht berechnet.

C&R-Baum, CHAID, Quest

Beim Erstellen von Modellen für große Datenmengen und -vielfalt wird das Ziel normalerweise in "Sehr große Datasets" geändert oder es werden Aufteilungen angegeben.

- Fortlaufendes Training vorhandener PSM-Modelle wird nicht unterstützt.
- Das Modellerstellungsziel Standard wird nur empfohlen, wenn Aufteilungsfelder so definiert sind, dass die Anzahl an Datensätzen in den einzelnen Aufteilungen nicht "zu groß" ist, wobei die Definition von "zu groß" von der Leistungsstärke einzelner Knoten in Ihrem Hadoop-Cluster abhängt. Im Gegensatz dazu müssen Sie auch darauf bedacht sein, sicherzustellen, dass Aufteilungen nicht so fein definiert sind, dass zu wenige Datensätze für die Erstellung eines Modells vorhanden sind.
- Das Ziel Boosting wird nicht unterstützt.
- Das Ziel Bagging wird nicht unterstützt.
- Das Ziel Sehr große Datasets wird nicht empfohlen, wenn wenige Datensätze vorhanden sind. Oft wird dann entweder kein Modell oder ein vermindertes Modell erstellt.
- Interaktive Sitzungen werden nicht unterstützt.
- Die Genauigkeitsstatistik wird für aufgeteilte Modelle nicht berechnet.
- Wenn ein Aufteilungsfeld vorhanden ist, unterscheiden sich Baummodelle, die lokal in Modeler erstellt wurden, geringfügig von Baummodellen, die von Analytic Server erstellt wurden. Daher werden unterschiedliche Scores erzeugt. In beiden Fällen sind die Algorithmen gültig. Die von Analytic Server verwendeten Algorithmen sind einfach nur neuer. Da Baumalgorithmen zahlreiche heuristische Regeln aufweisen, ist der Unterschied zwischen den beiden Komponenten normal.

Modellscoring

Alle für die Modellierung unterstützten Modelle werden auch für das Scoring unterstützt. Außerdem werden lokal erstellte Modellnuggets für die folgenden Knoten für das Scoring unterstützt: C&RT,

Quest, CHAID, Linear, Neuronales Netz (unabhängig davon, ob es ein Standard-, Boosting- oder Bagging-Modell oder ein Modell für sehr umfangreiche Datasets ist), Regression, C5.0, Logistisch, Genlin, GLMM, Cox, SVM, Bayes-Netz, TwoStep, KNN, Entscheidungsliste, Diskriminanzanalyse, Selbstlernfunktion, Anomalieerkennung, Apriori, Carma, K-Means, Kohonen, R und Textmining.

- Raw Propensity und Adjusted Propensity werden nicht gescort. Als Problemumgehung können Sie denselben Effekt erzielen, indem Sie die Raw Propensity mithilfe eines Ableitungsknotens mit dem folgenden Ausdruck berechnen: if 'predicted-value' == 'value-of-interest' then 'prob-of-that-value' else 1-'prob-of-that-value' endif

R

Die R-Syntax im Nugget sollte aus Operationen bestehen, die jeweils nur für einen Datensatz ausgeführt werden.

Ausgabe

Die Knoten Matrix, Analyse, Data Audit, Transformieren, Globalwerte, Statistik, Mittelwert und Tabelle werden unterstützt. Weitere Hinweise zur Funktionalität von unterstützten Knoten folgen.

Data Audit

Der Knoten Data Audit kann den Modus für stetige Felder nicht erzeugen.

Mittelwert

Der Knoten Mittelwert kann zu einem Standardfehler oder einem 95%-Konfidenzintervall führen.

Tabelle

Der Tabellenknoten wird unterstützt, indem eine temporäre Analytic Server-Datenquelle geschrieben wird, die die Ergebnisse der vorgeordneten Operationen enthält. Der Knoten Tabelle blättert dann durch den Inhalt der Datenquelle.

Exportieren

Ein Datenstrom kann mit einem Analytic Server-Quellenknoten beginnen und mit einem anderen Exportknoten als dem Analytic Server-Exportknoten enden. Die Daten werden jedoch von HDFS in SPSS Modeler Server und schließlich an die Exportposition verschoben.

Pushback in HCatalog/Hive

Pushback in HCatalog/Hive

Beim Arbeiten mit Daten in einer partitionierten Hive-Tabelle können Sie Ihren Modeler-Datenstrom so konfigurieren, dass für die Auswahl der gewünschten Partitionen ein Pushback an Hive durchgeführt wird.

1. Beginnen Sie Ihren Datenstrom mit einem Analytic Server-Quellenknoten, der die HCatalog/Hive-Datenquelle referenziert.
2. Stellen Sie wie gehabt Verbindungen zu anderen Knoten her.

In den folgenden Abschnitten werden Analytic Server-Operationen und -Funktionen bereitgestellt, die die SQL-Generierung unterstützen.

Analytic Server-Operationen

Tabelle 1. Datensatzoperationen

IBM SPSS Modeler-Operationen, die die SQL-Generierung unterstützen		IBM SPSS Analytic Server-Operationen	BigSQL	Hive
Wählen				Ja*
Beispiel			Ja*	Ja*
Sortieren			Ja	Ja
Balancieren				
Duplikat			Ja	Ja

Tabelle 1. Datensatzoperationen (Forts.)				
IBM SPSS Modeler-Operationen, die die SQL-Generierung unterstützen		**IBM SPSS Analytic Server-Operationen**	**BigSQL**	**Hive**
Aggregieren	SUMME		Ja	Ja
	MITTELWERT		Ja	Ja
	MIN		Ja	Ja
	MAX		Ja	Ja
	SDEV		Ja	Ja
	ANZAHL		Ja	Ja
	COUNTNON		Ja	Ja
	VARIANCE		Ja	Ja
	@	secondlargest	Nein	
	@	thirdlargest	Nein	
	@	sumOfSquare		
RFM-Aggregat				Nein
Zusammen-führen	inner		Ja	Ja
	outer		Ja	Ja
	anti		Nein	Nein
	@	join.cartesian	Nein	
Anhängen			Ja	Ja
Streaming-ZR			Nein	Nein
Streaming-TCM			Nein	Nein

@ - In SPSS Modeler gibt es keine entsprechende Funktion.

* - Der Knoten unterstützt viele Funktionen, von denen einige ein Pushback unterstützen.

Tabelle 2. Feldoperationen			
IBM SPSS Modeler-Operationen, die die SQL-Generierung unterstützen	**IBM SPSS Analytic Server-Operationen**	**BigSQL-Push-back**	**Hive-Pushback**
Typ			Ja
Filter		Ja	Ja
Ableiten		Ja*	Ja*
Füller			Ja
Umcodieren		Ja	Ja
Klassierung			Nein
RFM-Analyse			Nein
Ensemble			Nein

Tabelle 2. Feldoperationen (Forts.)			
IBM SPSS Modeler-Operationen, die die SQL-Generierung unterstützen	**IBM SPSS Analytic Server-Operationen**	**BigSQL-Push-back**	**Hive-Pushback**
Partitionierung			Nein
Dichotom			Ja
Umstrukturieren			Nein
Felder ordnen			Ja
Reprojizieren			Nein
Zeitintervalle			Nein

* - Der Knoten unterstützt viele Funktionen, von denen einige ein Pushback unterstützen.

Tabelle 3. Exportoperationen			
IBM SPSS Modeler-Operationen, die die SQL-Generierung unterstützen	**IBM SPSS Analytic Server-Operationen**	**BigSQL-Push-back**	**Hive-Pushback**
DS			

Tabelle 4. Diagrammoperationen			
IBM SPSS Modeler-Operationen, die die SQL-Generierung unterstützen	**IBM SPSS Analytic Server-Operationen**	**BigSQL-Push-back**	**Hive-Pushback**
Diagrammtafel			Ja*
Diagramm			Nein
Multiplot			Nein
Zeitdiagramm			Nein
Verteilung			Ja
Histogramm			Nein
Sammlung			Nein
Netzdiagramm			Ja
Auswertung			Nein
Kartenvisualisierung			Nein
E-Plot			Nein

* - Der Knoten unterstützt viele Funktionen, von denen einige ein Pushback unterstützen.

Tabelle 5. Nuggetoperationen			
IBM SPSS Modeler-Operationen, die die SQL-Generierung unterstützen	**IBM SPSS Analytic Server-Operationen**	**BigSQL-Push-back**	**Hive-Pushback**
GLE			Nein
Linear-AS			Nein
LSVM			Nein
Tree-AS			Nein

Tabelle 5. Nuggetoperationen (Forts.)			
IBM SPSS Modeler-Operationen, die die SQL-Generierung unterstützen	**IBM SPSS Analytic Server-Operationen**	**BigSQL-Push-back**	**Hive-Pushback**
Zeitreihe			Nein
TCM			Nein
STP			Nein
TwoStep-AS			Nein
Assoziationsregeln			Nein

Tabelle 6. Ausgabeoperationen			
IBM SPSS Modeler-Operationen, die die SQL-Generierung unterstützen	**IBM SPSS Analytic Server-Operationen**	**BigSQL-Push-back**	**Hive-Pushback**
Tabelle			Nein
Matrix			Ja
Analyse			Nein
Data Audit			Nein
Transformieren			Ja
Statistik			Nein
Mittelwert			Nein
Bericht			Ja*
Globalwerte			Ja

* - Der Knoten unterstützt viele Funktionen, von denen einige ein Pushback unterstützen.

Analytic Server-Funktionen

Tabelle 7. Arithmetische Funktionen			
Arithmetische Funktionen von IBM SPSS Modeler	**Arithmetische Funktionen von IBM SPSS Analytic Server**	**BigSQL-Push-back**	**Hive-Pushback**
+	+	Ja	Ja
-	-	Ja	Ja
*	*	Ja	Ja
/	/	Ja	Ja
@	%	Ja	
@	^	Ja	
@	isNull	Ja	

@ - In SPSS Modeler gibt es keine entsprechende Funktion.

Tabelle 8. Bitfunktionen

IBM SPSS Modeler-Bitfunktionen	IBM SPSS Analytic Server-Bitfunktionen	BigSQL-Push-back	Hive-Pushback
@	bitAnd	Ja	
@	bitAndEqualZero	Ja	
@	bitAndNot	Ja	
@	bitAndNotEqualZero	Ja	
@	bitNot	Ja	
@	bitOr	Ja	
@	bitXor	Ja	
integer_bitcount	intBitCount	Nein	Nein
integer_leastbit	intLeastBit	Nein	Nein
integer_length	intLength	Nein	Nein
testbit	testBit	Nein	Nein

@ - In SPSS Modeler gibt es keine entsprechende Funktion.

Tabelle 9. Zeichenfunktionen

IBM SPSS Modeler-Zeichenfunktionen	IBM SPSS Analytic Server-Zeichenfunktionen	BigSQL-Push-back	Hive-Pushback
isAlphaCode	isAlphaCode	Ja	
isNumberCode	isNumberCode	Ja	
isLowerCode	isLowerCode	Ja	
isUpperCode	isUpperCode	Ja	
unicode_char	unicodeChar	Nein	Nein
unicode_value	unicodeValue	Nein	

Tabelle 10. Vergleichsfunktionen

IBM SPSS Modeler-Vergleichsfunktionen	IBM SPSS Analytic Server-Vergleichsfunktionen	BigSQL-Push-back	Hive-Pushback
=	==	Ja	Ja
>=	>=	Ja	Ja
>	>	Ja	Ja
<=	<=	Ja	Ja
<	<	Ja	Ja
/=	!=	Ja	Ja
keine	!	Ja	Ja
und	und	Ja	Ja

Tabelle 10. Vergleichsfunktionen (Forts.)

IBM SPSS Modeler-Vergleichsfunktionen	IBM SPSS Analytic Server-Vergleichsfunktionen	BigSQL-Pushback	Hive-Pushback
oder	oder	Ja	Ja
count_equal	countEqual	Ja	Ja
count_greater_than	countGreaterThan	Ja	Ja
count_less_than	countLessThan	Ja	Ja
count_not_equal	countNotEqual	Ja	Ja
count_nulls	countnulls	Ja	Ja
first_index	firstindex	Nein	Nein
first_non_null_index	firstnonnullindex	Ja	Ja
last_index	lastindex	Nein	Nein
last_non_null_index	lastnonnullindex	Ja	Ja
value_at	valueAt	Ja	Ja
@	hash	Nein	

@ - In SPSS Modeler gibt es keine entsprechende Funktion.

Tabelle 11. Konvertierungsfunktionen

IBM SPSS Modeler-Konvertierungsfunktionen	IBM SPSS Analytic Server-Konvertierungsfunktionen	BigSQL-Pushback	Hive-Pushback
to_Int	toint	Ja	Ja
to_real	toReal	Ja	Ja
to_string	toString	Nein	Ja
to_date	toDate	Ja	Ja
to_time	toTime	Nein	Nein
to_timestamp	toTimestamp	Nein	Nein

Tabelle 12. Datums- und Zeitfunktionen

Datums- und Zeitfunktionen von IBM SPSS Modeler	Datums- und Zeitfunktionen von IBM SPSS Analytic Server	BigSQL-Pushback	Hive-Pushback
datetime_now	now	Ja	Ja
Today	today	Nein	Nein
@	Format	Nein	Nein
datetime_date_name(DAY)	dayName	Nein	Nein
datetime_month_name(MONTH)	monthName	Nein	Nein
datetime_date(item)	date(Number)	Ja	Nein

Tabelle 12. Datums- und Zeitfunktionen (Forts.)

Datums- und Zeitfunktionen von IBM SPSS Modeler	Datums- und Zeitfunktionen von IBM SPSS Analytic Server	BigSQL-Pushback	Hive-Pushback
datetime_date(item)	date(timestamp)	Ja	Nein
@	date(string,format,locale)	Nein	
datetime_date(year,month,day)	date(year,month,day)	Ja	Ja
datetime_date(item)	time(Number)	Nein	Nein
datetime_date(item)	time(timestamp)	Ja	Nein
@	time(string,format,locale)	Nein	Nein
datetime_time(hour,minute,second)	time(hour,minute,second)		
	timefromtimestamp		Ja
datetime_timestamp(item)	timestamp(number)	Nein	Nein
datetime_timestamp(date,time)	timestamp(date,time)	Ja	Nein
@	timestamp(string,format,locale)	Nein	Nein
datetime_timestamp(year,month,day,hour,minute,second)	timestamp(y,m,d,h,m,s)		Nein
datetime_year(DATE)	Jahr	Ja	Nein
datetime_month(DATE)	Monat	Ja	Nein
datetime_day(DATE)	day	Ja	Nein
datetime_weekday(DATE)	dayOfWeek	Ja	Nein
datetime_hour(TIME)	hours	Ja	Nein
datetime_minute(TIME)	minutes	Ja	Nein
datetime_second(TIME)	seconds	Ja	Nein
@	milliseconds	Ja	
date_before	dateBefore	Ja	Nein
time_before	timeBefore	Ja	Nein
date_days_difference	daysDifference	Ja	Nein
date_weeks_difference	weeksDifference	Ja	Nein
date_months_difference	monthsDifference	Ja	Nein
date_years_difference	yearsDifference	Ja	Nein
time_hours_difference	hoursDifference	Ja	Nein
time_mins_difference	minutesDifference	Ja	Nein
time_secs_difference	secondsDifference	Ja	Nein

Tabelle 12. Datums- und Zeitfunktionen (Forts.)			
Datums- und Zeitfunktionen von IBM SPSS Modeler	**Datums- und Zeitfunktionen von IBM SPSS Analytic Server**	**BigSQL-Pushback**	**Hive-Pushback**
time_in_hours	timeInHours	Ja	Nein
time_in_seconds	timeInSeconds	Ja	Nein
date_in_days	dateInDays	Ja	Nein
date_in_weeks	dateInWeeks	Ja	Nein
date_in_months	dateInMonths	Ja	Nein
date_in_years	dateInYears	Ja	Nein

@ - In SPSS Modeler gibt es keine entsprechende Funktion.

Tabelle 13. Listenfunktionen			
IBM SPSS Modeler-Listenfunktionen	**IBM SPSS Analytic Server-Listenfunktionen**	**BigSQL-Pushback**	**Hive-Pushback**
@	argMax		
@	argMin		
@	concatList		
@	Gruppe		
@	indexMax		
@	indexMin		
@	map		
@	Trennwand		
@	reduceLeft		
@	reverse		
@	sortList		

@ - In SPSS Modeler gibt es keine entsprechende Funktion.

Tabelle 14. Informationsfunktionen			
IBM SPSS Modeler-Informationsfunktionen	**IBM SPSS Analytic Server-Informationsfunktionen**	**BigSQL-Pushback**	**Hive-Pushback**
is_date	isDate	Ja	Ja
is_datetime	isDateTime	Ja	Ja
is_integer	isInt	Ja	Ja
is_number	isNumber		Nein
is_real	isReal	Ja	Ja
is_string	isString		Nein

Tabelle 14. Informationsfunktionen (Forts.)

IBM SPSS Modeler-Informationsfunktionen	IBM SPSS Analytic Server-Informationsfunktionen	BigSQL-Pushback	Hive-Pushback
is_time	isTime	Ja	Ja
is_timestamp	isTimestamp	Ja	Ja

Tabelle 15. Mathematische Funktionen

Mathematische Funktionen von IBM SPSS Modeler	Mathematische Funktionen von IBM SPSS Analytic Server	BigSQL-Pushback	Hive-Pushback
abs	abs	Ja	Ja
div	divWhole	Nein	Nein
@	e	Ja	
exp	exp	Ja	Ja
fracof	frac	Ja	Ja
@	ln	Ja	
log10	log10	Ja	Ja
pi	pi	Ja	Ja
random	random	Ja	Ja
@	realToInt	Ja	
round	round	Ja	Ja
sign	sign	Ja	Ja
sqrt	sqrt	Ja	Ja

@ - In SPSS Modeler gibt es keine entsprechende Funktion.

Tabelle 16. Wahrscheinlichkeitsfunktionen

IBM SPSS Modeler-Wahrscheinlichkeitsfunktionen	IBM SPSS Analytic Server-Wahrscheinlichkeitsfunktionen	BigSQL-Pushback	Hive-Pushback
cdf_chisquare	cdfChiSquare	Nein	Nein
cdf_f	cdfF	Nein	Nein
cdf_normal	cdfNormal	Nein	Nein
cdf_t	cdfT	Nein	Nein

Tabelle 17. Statistische Funktionen

Statistische Funktionen von IBM SPSS Modeler	Statistische Funktionen von IBM SPSS Analytic Server	BigSQL-Pushback	Hive-Pushback
max_n	maxN	Ja	Ja

Tabelle 17. Statistische Funktionen (Forts.)

Statistische Funktionen von IBM SPSS Modeler	Statistische Funktionen von IBM SPSS Analytic Server	BigSQL-Pushback	Hive-Pushback
mean_n	meanN	Ja	Ja
min_n	minN	Ja	Ja
sdev_n	sDevN	Nein	Nein
sum_n	sumN	Nein	Nein

Tabelle 18. Trigonometrische Funktionen

Trigonometrische Funktionen von IBM SPSS Modeler	Trigonometrische Funktionen von IBM SPSS Analytic Server	BigSQL-Pushback	Hive-Pushback
sin	sin	Ja	Ja
sinh	sinh	Ja	Ja
cos	cos	Ja	Ja
cosh	cosh	Ja	Ja
tan	tan	Ja	Ja
tanh	tanh	Ja	Ja
arcsin	arcsin	Ja	Ja
arcsinh	arcsinh	Ja	Ja
arccos	arccos	Ja	Ja
arccosh	arccosh	Ja	Ja
arctan	arctan	Ja	Ja
arctan2	arctan2	Ja	Nein
arctanh	arctanh	Ja	Ja
@	toDegrees	Ja	
@	toRadians	Ja	

@ - In SPSS Modeler gibt es keine entsprechende Funktion.

Tabelle 19. Zeichenfolgefunktionen

IBM SPSS Modeler-Zeichenfolgefunktionen	IBM SPSS Analytic Server-Zeichenfolgefunktionen	BigSQL-Pushback	Hive-Pushback
allbutfirst	allButFirst	Ja	Ja
allbutlast	allButLast	Ja	Ja
alphabefore	alphaBefore	Ja	Ja
@	charAt	Ja	
@	compare	Nein	

Tabelle 19. Zeichenfolgefunktionen (Forts.)

IBM SPSS Modeler-Zeichenfolgefunktionen	IBM SPSS Analytic Server-Zeichenfolgefunktionen	BigSQL-Pushback	Hive-Pushback
@	Zähler	Nein	
@	detectLanguage	Nein	
endstring	endString	Ja	Ja
isstartstring	isStartString	Ja	Ja
isendstring	isEndString	Ja	Ja
ismidstring	isMidString	Nein	Ja
issubstring	isSubString	Ja	Ja
issubstring_count	isSubStringCount	Nein	Nein
issubstring_lim	isSubStringLim	Nein	Nein
@	jsonPath	Nein	
@	concat	Ja	
last	last	Ja	Ja
Länge	Länge	Ja	Ja
@	lengthInBytes	Ja	Ja
locchar	locChar	Ja	
locchar_back	locCharBack	Ja	Ja
uppertolower	lower	Ja	
trimstart	lTrim	Ja	Ja
ersetzen	ersetzen	Ja	Nein
replicate	replicate	Ja	Ja
trimend	rTrim	Ja	Ja
skipchar	skipChar	Nein	Nein
skipchar_back	skipCharBack	Nein	Nein
soundex	soundEx	Ja	Ja
soundex_difference	soundExDifference	Ja	Ja
startstring	startString	Ja	Ja
stripchar	stripChar	Ja	
strmember	strMember	Ja	
substring	subString	Ja	Nein
substring_between	subStringBetween	Ja	Nein
trim	trim	Ja	Ja
lowertoupper	upper	Ja	

@ - In SPSS Modeler gibt es keine entsprechende Funktion.

Tabelle 20. Georäumliche Funktionen			
Georäumliche Funktionen von IBM SPSS Modeler	**Georäumliche Funktionen von IBM SPSS Analytic Server**	**BigSQL-Pushback**	**Hive-Pushback**
close_to	closeTo	Nein	Nein
crosses	crosses	Nein	Nein
@	intersect	Nein	
@	touch	Nein	
overlap	overlap	Nein	Nein
within	within	Nein	Nein
@	contain	Nein	
@	northOf	Nein	
@	southOf	Nein	
@	eastOf	Nein	
@	westOf	Nein	
@	centroid	Nein	
area	area	Nein	Nein
num_points	numPoints	Nein	Nein

@ - In SPSS Modeler gibt es keine entsprechende Funktion.

Kapitel 3. Fehlerbehebung

In diesem Abschnitt werden einige allgemeine Probleme bei der Verwendung sowie Wege zu deren Lösung beschrieben.

Datenquellen

Für partitionierte Spalten in HCatalog-Datenquellen definierte Filter werden nicht berücksichtigt.
Dieses Problem tritt in einigen Versionen von Hive auf und kann in folgenden Situationen auftreten.

- Wenn Sie eine HCatalog-Datenquelle definieren und einen Filter in der Datenquellendefinition angeben.
- Wenn Sie einen Modeler-Datenstrom mit einem Filterknoten erstellen, der auf die partitionierte Tabellenspalte verweist.

Als Problemumgehung kann dem Modeler-Datenstrom ein Ableitungsknoten hinzugefügt werden, der ein neues Feld erstellt, dessen Werte der partitionierten Spalte entsprechen. Der Filterknoten sollte auf dieses neue Feld verweisen.

Oracle NoSQL

Beim Herstellen einer Verbindung zu einer Oracle NoSQL-Datenquelle treten Fehler "Execution failed" auf.
Das Problem ist das Ergebnis des veralteten Speicherhandlers HiveKVStorageHandler.jar. Sie müssen einen aktualisierten Speicherhandler verwenden. Sie finden die aktualisierte Datei unter https://github.com/dvasilen/HiveKVStorageHandler3/raw/HADOOP_2.6-HIVE-1.2.0-KV-3.3.4/release/hive-kv-storage-handler-1.2.0-3.3.4.jar.

```
hive-kv-storage-handler-1.2.0-3.3.4.jar
```

1. Kopieren Sie die JAR-Datei in das Verzeichnis Hive `{HIVE_HOME}/auxlib` und das Verzeichnis Analytic Server `{AS_ROOT}/ae_wlpserver/usr/servers/aeserver/apps/AE_BOOT.war/WEB-INF/lib` .
2. Führen Sie `{AS_ROOT}/bin/hdfsUpdate.sh` aus, um die Änderungen an HDFSweiterzugeben.
3. Starten Sie Analytic Server erneut, damit die Änderungen wirksam werden.

Anmerkung: Die Speicherhandlerklasse `oracle.kv.hadoop.hive.table.TableStorageHandler` wird empfohlen, wenn die Datenbank "Oracle NoSQL 3.0" verwendet wird. Bei dieser Klasse müssen Benutzer Daten mit einer Tabellenmetapher organisieren.

Bemerkungen

Die vorliegenden Informationen wurden für Produkte und Services entwickelt, die auf dem deutschen Markt angeboten werden. IBM stellt dieses Material möglicherweise auch in anderen Sprachen zur Verfügung. Für den Zugriff auf das Material in einer anderen Sprache kann eine Kopie des Produkts oder der Produktversion in der jeweiligen Sprache erforderlich sein.

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 dieser Dokumentation 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
USA

For license inquiries regarding double-byte (DBCS) information, contact the IBM Intellectual Property Department in your country or send inquiries, in writing, to:

Intellectual Property Licensing
Legal and Intellectual Property Law
IBM Japan Ltd.
19-21, Nihonbashi-Hakozakicho, Chuo-ku
Tokyo 103-8510, Japan

INTERNATIONAL BUSINESS MACHINES CORPORATION PROVIDES THIS PUBLICATION "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some jurisdictions do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

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

Die angeführten Leistungsdaten und Kundenbeispiele dienen nur zur Illustration. Die tatsächlichen Ergebnisse beim Leistungsverhalten sind abhängig von der jeweiligen Konfiguration und den Betriebsbedingungen.

Alle Informationen zu Produkten anderer Anbieter stammen vonIBM den Anbietern der aufgeführten Produkte, deren veröffentlichten Ankündigungen oder anderen allgemein verfügbaren Quellen. IBM hat diese Produkte nicht getestet und kann die Genauigkeit der Leistung, der Kompatibilität oder anderer Ansprüche im Zusammenhang mit Nicht-IBM-Produkten nicht bestätigen. Fragen zu den Leistungsmerkmalen von Produkten anderer Anbieter als IBM 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.

Alle von IBM angegebenen Preise sind empfohlene Richtpreise und können jederzeit ohne weitere Mitteilung geändert werden. Händlerpreise können unter Umständen von den hier genannten Preisen abweichen.

Diese Veröffentlichung dient nur zu Planungszwecken. Die in dieser Veröffentlichung enthaltenen Informationen können geändert werden, bevor die beschriebenen Produkte verfügbar sind.

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.

COPYRIGHTLIZENZ:

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.

Kopien oder Teile der Beispielprogramme bzw. daraus abgeleiteter Code müssen folgenden Copyrightvermerk beinhalten:

Marken

IBM, das IBM Logo und ibm.com sind Marken oder eingetragene Marken der IBM Corporation in vielen Ländern weltweit registriert. 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.

Adobe, das Adobe Logo, PostScript und das PostScript Logo sind entweder eingetragene Marken oder Marken von Adobe Systems Incorporated in den USA und/oder anderen Ländern.

IT Infrastructure Library ist eine eingetragene Marke der Central Computer and Telecommunications Agency. Die Central Computer and Telecommunications Agency ist nunmehr in das Office of Government Commerce eingegliedert worden.

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 ihrer Tochtergesellschaften in den USA oder anderen Ländern.

Linux ist eine eingetragene Marke von Linus Torvalds in den USA und/oder anderen Ländern.

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

ITIL ist eine eingetragene Marke und eine eingetragene Gemeinschaftsmarke von The Minister for the Cabinet Office, und ist im US- Patent and Trademark Office eingetragen ist.

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

Cell Broadband Engine wird unter Lizenz verwendet und ist eine Marke der Sony Computer Entertainment, Inc. in den USA und/oder anderen Ländern.

Linear Tape-Open, LTO, das LTO-Logo, Ultrium und das Ultrium-Logo sind Marken von HP, der IBM Corporation und von Quantum in den USA und/oder anderen Ländern.