

IBM SPSS Analytic Server
Versione 3.5

Guida per l'utente



Nota

Prima di utilizzare queste informazioni e il prodotto che supportano, leggere le informazioni in [“Informazioni particolari” a pagina 43](#).

Informazioni sui prodotti

Questa edizione si applica alla versione 3, release 5, modifica 0 di IBM® SPSS Analytic Server e a tutte le successive release e modifiche fino a quando diversamente indicato nelle nuove edizioni.

© Copyright International Business Machines Corporation .

Indice

- Capitolo 1. Server analitico Console..... 1**
 - Origini dati 1
 - Impostazioni (origini dati file)..... 6
 - Associazioni Campo HCatalog..... 12
 - Utilizzo delle origini dati HCatalog..... 13
 - Anteprima e Metadati (origini dati)..... 19
 - Progetti 19
 - Gestione degli utenti e dei gruppi 21
 - Gestione dei ruoli utente e di gruppo 22
 - Regole di denominazione 23
- Capitolo 2. SPSS Modeler Integrazione.....25**
 - Nodi supportati..... 25
 - Pushback a HCatalog/Hive..... 29
- Capitolo 3. Risoluzione dei problemi..... 41**
- Informazioni particolari..... 43**
 - Marchi..... 44

Capitolo 1. Server analitico Console

Server analitico fornisce un'interfaccia client sottile per la gestione delle origini dati e dei progetti.

Accesso

1. Inserire l'URL della Server analitico nella barra degli indirizzi del browser. L'URL può essere ottenuto dall'amministratore del server.
2. Immettere il nome utente con il quale accedere al server.
3. Immettere la password associata al nome utente specificato.

Nota: Il nome utente inserito durante il prompt di login della console Server analitico viene immesso senza il suffisso del nome del dominio. Di conseguenza, quando vengono definiti più - realms, gli utenti vengono presentati con un elenco a discesa **Realms** che consente loro di selezionare il realm appropriato. Quando viene definito un solo realm, gli utenti non vengono presentati con un elenco a discesa **Realms** quando si firma in Server analitico.

Dopo il login, viene visualizzata la home Console.

Navigazione nella console

- L'intestazione visualizza il nome del prodotto, il nome dell'utente attualmente collegato e il link al sistema di guida. Il nome dell'utente attualmente collegato è il capo di un elenco a discesa che include il collegamento logout.
- L'area dei contenuti visualizza le azioni che puoi prendere dalla casa Console.

Origini dati

Un'origine dati è una raccolta di record, più un modello di dati, che definiscono un dataset per l'analisi. La fonte dei record può essere un file (testo delimitato, testo a larghezza fissa, Excel) su HDFS, un database relazionale, HCatalog o geospaziale. Il modello di dati definisce tutti i metadati (nomi di campo, storage, livello di misurazione e così via) necessari per l'analisi dei dati. I proprietari di origine dati possono concedere o limitare l'accesso alle fonti di dati.

Elenco delle origini dati

La pagina principale delle origini dati fornisce un elenco di fonti di dati di cui l'utente corrente è membro.

- Clicca su un nome di origine dati per visualizzare i suoi dettagli e modificarne le proprietà.
- Digitare nell'area di ricerca per filtrare l'elenco per visualizzare solo le origini dati con la stringa di ricerca nel loro nome.
- Clicca su **Nuovo** per creare una nuova origine dati con il nome e il tipo di contenuto specificati nella finestra di dialogo **Aggiungi nuova origine dati**.
 - Consultare [“Regole di denominazione”](#) a [pagina 23](#) per limitazioni sui nomi che è possibile dare alle origini dati.
 - I tipi di contenuto disponibili sono File, Database, HCatalog e Geospatial.

Note:

- L'opzione HCatalog è disponibile solo se Analytic Server è stato configurato per funzionare con tali origini dati.
- Il tipo di HCatalog non è disponibile per HDP 3.0 o versioni successive e CDH 6.0 o versioni successive.
- Il tipo di contenuto non può essere modificato una volta selezionato.

- È possibile importare / esportare più fonti di dati in un'unica azione.
- Clicca su **Elimina** per rimuovere l'origine dati. Questa azione lascia intatti tutti i file associati alla fonte dati.
- Clicca su **Aggiorna** per aggiornare la quotazione.
- L'elenco a discesa **Azioni** esegue l'azione selezionata.
 1. Selezionare **Esporta** per creare un archivio dalle origini dati selezionate e salvare l'archivio sul file system locale. L'archivio include eventuali file aggiunti alle origini dati selezionate in modalità **Progetti** o **Data source**.
 - Nota:** Quando viene selezionata una sola origine dati, il nome del file di archivio condivide il nome origine dati selezionato. Se viene selezionata più di una fonte dati, il nome del file di archivio viene impostato per default sul nome `datasources.zip`.
 2. Selezionare **Importa** per importare gli archivi creati dall'azione Export.
 - Nota:** I file di archivio che contengono informazioni provenienti da più origini dati non possono essere importati. In questi casi, i singoli archivi di origine dati devono prima essere estratti dall'archivio `datasources.zip`.
 3. Selezionare **Duplicato** per creare una copia dell'origine dati.

Dettagli di origine dati individuali

L'area dei contenuti è suddivisa in diverse sezioni, che possono dipendere dal tipo di contenuto dell'origine dati.

Dettagli

Queste impostazioni sono comuni a tutti i tipi di contenuto.

Nome

Un campo di testo modificabile che mostra il nome dell'origine dati.

Nome visualizzato

Un campo di testo modificabile che mostra il nome dell'origine dati come visualizzato in altre applicazioni. Se questo è vuoto, il Nome viene utilizzato come nome di visualizzazione.

Descrizione

Un campo di testo modificabile per fornire un testo esplicativo sull'origine dati.

È pubblico

Una casella di controllo che indica se qualcuno può vedere l'origine dati (controllata) o se gli utenti e i gruppi devono essere esplicitamente aggiunti come membri (sdoganati).

È condivisione globale

Una casella di controllo che controlla se la RDD di Spark è memorizzata nella cache globale. Quando selezionato, il RDD di Spark viene sempre memorizzato nella cache globale. Quando deselezionato, il RDD di Spark viene rimosso dalla cache globale quando nessun lavoro Spark lo sta utilizzando.

Attributi personalizzati

Le applicazioni possono allegare proprietà a fonti di dati, ad esempio se l'origine dati è temporanea, attraverso l'uso di attributi personalizzati. Questi attributi sono esposti nella console Server analitico per fornire ulteriori approfondisci su come le applicazioni utilizzano l'origine dati.

Clicca su **Salva** per mantenere lo stato attuale delle impostazioni.

Condivisione

Queste impostazioni comuni a tutti i tipi di contenuto.

È possibile condividere la proprietà di un'origine dati aggiungendo utenti e gruppi come Autori o Readers.

- Digitando nella casella di testo filtri su utenti e gruppi con la stringa di ricerca nel loro nome. Selezionare **Autore** o **Reader** dall'elenco a discesa per assegnare il loro ruolo all'interno dell'origine dati. Clicca su **Aggiungi membro** per aggiungerli all'elenco dei membri.
- Per rimuovere un partecipante, selezionare un utente o un gruppo nell'elenco dei membri e fare clic su **Rimuovi membro**.

Nota: Gli utenti con il ruolo **Amministratore** hanno letto e scrivono l'accesso ad ogni fonte di dati, indipendentemente dal fatto che siano specificamente elencati come membro.

Input file

Impostazioni specifiche per definire le origini dati con tipo di contenuto file.

File Viewer

Mostra i file disponibili per l'inclusione nella fonte dati. Selezionare la modalità **Progetti** per visualizzare i file all'interno della struttura del progetto Server analitico, **Data source** per visualizzare i file memorizzati all'interno di un'origine dati o **File system** per visualizzare il file system (tipicamente HDFS). È possibile sfogliare la struttura delle cartelle, ma HDFS non è modificabile affatto e in modalità **Progetti** non è possibile aggiungere file, creare cartelle o cancellare elementi a livello di root, ma solo all'interno di progetti definiti. Per creare, modificare o eliminare un progetto, utilizzare [Progetti](#).

- Clicca su **Carica** per caricare un file nella sorgente dati corrente o su project/subfolder. È possibile sfogliare e selezionare più file in una singola directory.

Nota: I file vengono caricati sul file system distribuito. È possibile trovare i file caricati nella struttura della directory `/analytic-root`, sotto l'apposito tenant, data source o progetto (a seconda della modalità scelta) e sottocartella. Ad esempio, se si:

1. Accedere al tenant `ibm`
2. Creare un'origine dati denominata `fraudDetection`
3. Selezionare la modalità **Data source**
4. Creare una sottocartella chiamata `historicalData`
5. Carica un file `charges2015.csv`

Poi il file può essere trovato sul file system distribuito in `/analytic-root/ibm/.datasource/fraudDetection/historicalData/charges2015.csv`. Se, invece, lei:

1. Accedere al tenant `ibm`
2. Creare un'origine dati denominata `fraudDetection`
3. Selezionare la modalità **Progetto**
4. Selezionare un progetto esistente denominato `creditProcessing`
5. Creare una sottocartella chiamata `historicalData`
6. Carica un file `charges2015.csv`

Poi il file può essere trovato sul file system distribuito in `/analytic-root/ibm/creditProcessing/historicalData/charges2015.csv`.

- Fare clic su **Nuova cartella** per creare una nuova cartella sotto la cartella corrente, con il nome specificato nella finestra di dialogo Nuovo Nome cartella.
- Clicca su **Scarica** per scaricare i file selezionati sul file system locale.
- Clicca su **Elimina** per rimuovere i file / cartelle selezionati.

File inclusi nella definizione dell'origine dati

Utilizzare il pulsante di spostamento per aggiungere file e cartelle selezionate a, o rimuoverli, l'origine dati. Per ogni file o cartella selezionata nell'origine dati, fare clic su [Impostazioni](#) per definire le specifiche per la lettura del file.

Quando i file multipli sono inclusi in un'origine dati, devono condividere un metadati comune; cioè ogni file deve avere lo stesso numero di campi, i campi devono essere analizzati nello stesso ordine in ogni file e ogni campo deve avere la stessa memoria su tutti i file. I mismatch tra i file possono causare la mancata creazione della console per creare i valori Anteprima e Metadati, o comunque valori validi da analizzare come non validi (null) quando Server analitico legge il file.

Elezioni del database

Specificare i parametri di connessione per il database che contiene il contenuto record.

Database

Selezionare il tipo di database a cui connettersi. Scegli tra: Db2, Greenplum, Apache Impala, Amazon Redshift, MySQL, Netezza, Oracle, SQL Server, TeraData, Hive, DashDB o BigSQL. Se il tipo che si sta cercando non è elencato, chiedere all'amministratore del server di configurare Server analitico con il driver JDBC appropriato.

Nota: Server analitico supporta i database MySQL ubicati su sistemi remoti.

Hive Tipo di connessione

Questa opzione è disponibile solo quando Hive è selezionato come tipo di **database**. Selezionare il tipo di connessione, **Server singolo** o **Alta disponibilità**. **Server singolo** viene utilizzato quando viene utilizzato un singolo server Hive; **Alta disponibilità (High Availability)** viene utilizzato quando viene utilizzato un cluster di server Hive altamente disponibile. Quando si seleziona **Alta disponibilità**, sono disponibili le seguenti opzioni:

Zookeeper Quorum

Immettere un elenco separato da virgole per tutti gli host: porte del server Zookeeper (ad esempio: zkhost1:2181, zkhost2:2181).

Namespace

Inserire lo spazio nome root Hive su Zookeeper. Ad esempio, hiveserver2 o hiveserver2-hive2 (quando hiveserver2 interattivo è abilitato e utilizzato su HDP 2.6).

Note:

- I Valori **Zookeeper Quorum** e **Name Space** si trovano nel file `hive-site.xml`.
- Per impostazione predefinita, Hive High Availability è disabilitato in Cloudera e deve essere abilitato manualmente.
- Quando si utilizza un'origine dati Hive in un ambiente nonKerberos, è necessario assicurarsi che Username immesso nella sezione **Selezioni database** sia uguale a `login AS user`.

Indirizzo del server

Inserire l'URL del server che ospita il database.

Porta server

Il numero di porta su cui il database ascolta.

Nome database

Il nome del database a cui si desidera collegarsi.

Nome utente

Se il database è protetto da password, inserire il proprio nome utente.

Password

Se il database è protetto da password, inserire la password.

Nome tabella

Inserire il nome di una tabella dal database che si desidera utilizzare.

Lettture simultane massime

Inserire il limite sul numero di query parallele che possono essere inviate da Server analitico al database da leggere dalla tabella specificata nell'origine dati.

Elezioni HCatalog

Specificare i parametri per accedere ai dati gestiti in Apache HCatalog.

Database

Il nome del database HCatalog.

Nome tabella

Inserire il nome di una tabella dal database che si desidera utilizzare.

Filtro

Il filtro di partizione per la tabella, se la tabella è stata creata come tabella partizionata. Il filtro HCatalog è supportato solo su chiavi di partizione Hive di tipo stringa.

Nota: Gli operatori `=`, `<>` e `LIKE` non sembrano funzionare in determinate distribuzioni Hadoop. Si tratta di un problema di compatibilità tra HCatalog e quelle distribuzioni.

Associazioni Campo HCatalog

Visualizza la mappatura di un elemento in HCatalog in un campo nell'origine dati. Clicca su [Modifica](#) per modificare le mappature del campo.

Nota: Dopo aver creato un'origine dati basata su HIC che espone i dati da una tabella Hive, è possibile scoprire che quando la tabella Hive è formata da un elevato numero di file di dati, vi è un ritardo sostanziale sostenuto ogni volta che Server analitico inizia a leggere i dati provenienti dall'origine dati. Se si notano tali ritardi, ricostruire la tabella Hive utilizzando un numero minore di file di dati di dimensioni maggiori e ridurre il numero di file a 400 o meno.

Elezioni geospaziali

Specificare i parametri per accedere ai dati geografici.

tipo geospaziale

I dati geografici possono provenire da un servizio di mappa online o da un file di forma.

Se si utilizza un servizio di mappa, specificare l'URL del servizio e selezionare lo strato di mappa che si desidera utilizzare.

Se si sta utilizzando un file di forma, selezionare o caricare il file shape. Da notare che un file di forma è in realtà una serie di file con un nomefile comune, memorizzati nella stessa directory. Selezionare il file con il suffisso SHP. Analytic Server cercherà e utilizzerà gli altri file. Due file aggiuntivi con i suffissi SHX e DBF devono essere sempre presenti; a seconda del file shape potrebbe essere presente anche una serie di file aggiuntivi.

Anteprima e Metadati

Dopo aver specificato le impostazioni per l'origine dati, fare clic su [Anteprima e Metadati](#) per verificare e confermare le specifiche dell'origine dati.

Output

Le origini dati con file, database o tipo di contenuto HCatalog possono essere accodate o sovrascritte dall'output da flussi eseguiti su Server analitico. Selezionare **Fai scrivibile** per abilitare l'appendendo o la sovrascrittura e:

- Per le origini dati con tipo di contenuto del database, scegliere una tabella database di output dove vengono scritti i dati di output.
- Per le origini dati con il tipo di contenuto dei file:

1. Scegli una cartella di output dove vengono scritti i nuovi file.

Suggerimento: Utilizzare una cartella separata per ogni sorgente dati così è più facile tenere traccia delle associazioni tra file e origini dati.

2. Selezionare un formato file; **CSV** (valore separato da virgola) o **Formato binario Splittabile**.
3. Opzionalmente selezionare **Fai file di sequenza**. Ciò è utile se si desidera creare file compressi suddivisibili utilizzabili nei lavori MapReduce downstream.
4. Selezionare **Newline possono essere scappate** se la tua emissione è CSV e hai campi di stringa che contengono newline o caratteri di ritorno a capo. Questo causerà 'ogni newline da scrivere come backslash seguito dalla lettera "n", il ritorno al carrello come backslash seguito dalla lettera "r", e la backslash come due backslashes consecutivi. Tali dati devono essere letti con la stessa impostazione. Suggeriamo fortemente di utilizzare il formato binario Splittabile quando si gestiscono i dati delle stringhe che contengono i caratteri di ritorno di newline o di trasporto.

5. Selezionare un formato di compressione. L'elenco include tutti i formati che sono stati configurati per l'utilizzo con la tua installazione di Server analitico.

Nota: Alcune combinazioni di formato di compressione e formato file risultano in un output che non può essere suddiviso e non è quindi adatto per un'ulteriore elaborazione MapReduce. Server analitico produce un avviso nella sezione Output quando si effettua tale selezione.

- Per le origini dati con tipo di contenuto HCatalog, scegliere una tabella hive in uscita dove vengono scritti i dati di output.

Note e limitazioni di origine dati HCatalog:

- La tabella di origine dati HCatalog deve esistere prima di lavorare con Server analitico (Server analitico non crea la tabella richiesta).
- Il modello di metadata/data della tabella deve essere coerente con il modello di dati dei risultati da esportare.
- L'origine dati HCatalog supporta solo la modalità di accodamento; la modalità di sostituzione non è supportata.

Impostazioni (origini dati file)

La finestra di dialogo Impostazioni consente di definire le specifiche per la lettura dei dati basati su file. Le impostazioni si applicano a tutti i file selezionati e tutti i file all'interno delle cartelle selezionate che corrispondono ai criteri presenti nella scheda **Cartella**. La specifica delle impostazioni del parser errate per un file può causare la mancata creazione della console per creare i valori Anteprima e Metadati, o comunque valori validi da analizzare come non validi (null) quando Server analitico legge il file.

Scheda Impostazioni

La scheda Impostazioni consente di specificare le impostazioni del tipo di file e del parser specifico per il tipo di file.

È possibile definire le origini dati utilizzando file compressi per qualsiasi formato file supportato. I formati di compressione supportati includono Gzip, Deflate, Bz2, Snappy e IBM CMX.

Tipo di file delimitato

I file delimitati sono file di testo a campo libero, i cui record contengono un numero costante di campi ma un numero variegato di caratteri per campo. I file delimitati tipicamente hanno *.csv o *.tab file estensioni. Per ulteriori informazioni, consultare [“Impostazioni tipo di file delimitato” a pagina 7](#).

Tipo di file fisso

I file di testo a campo fisso sono file i cui campi non sono delimitati ma iniziano nella stessa posizione e sono di lunghezza fissa. I file di testo in campo fisso hanno tipicamente una estensione file *.dat. Per ulteriori informazioni, consultare [“Impostazioni tipo di file fisso” a pagina 8](#).

Tipo di file semi strutturato

I file semi - strutturati (come *.log) sono file di testo che hanno una struttura prevedibile che può essere mappato a campi tramite espressioni regolari, ma non sono altamente strutturati come file delimitati. Per ulteriori informazioni, consultare [“Impostazioni tipo di file semi - strutturate” a pagina 9](#).

Tipo di file Analytics di testo

I file di Text Analytics sono documenti (come *.doc, *.pdf o *.txt) che possono essere analizzati utilizzando SPSS Text Analytics.

Saltare le linee vuote

Specifica se ignorare le righe vuote nel contenuto del testo estratto. Il default è **No**.

Separatore di riga

Specifica la stringa che definisce una nuova linea. Il default è il nuovo carattere di riga "\n".

Tipo di file SPSS Statistics

I file SPSS Statistics (*.sav, *.zsav) sono file binari che contengono un modello di dati. Non sono necessarie ulteriori impostazioni sulla scheda Impostazioni per questo tipo di file.

Tipo di file di formato binario splittabile

Specifica che il tipo di file è un file di formato binario splittabile (*.asbf). Questo tipo di file può rappresentare tutti i tipi di campo Server analitico (a differenza di CSV, che non possono rappresentare affatto i campi di elenco e richiede impostazioni speciali per gestire le newline incorporate e i ritorni di trasporto). Non sono necessarie ulteriori impostazioni sulla scheda Impostazioni per questo tipo di file.

Tipo di file di sequenza

I file di sequenza (*.seq) sono file di testo strutturati come coppie chiave / valore. Sono comunemente utilizzati come formato intermedio nei job MapReduce .

Tipo di file excel

Specifica che il tipo di file è un file Microsoft Excel (*.xls, *.xlsx). Per ulteriori informazioni, consultare ["Impostazioni tipo file excel"](#) a pagina 10.

Impostazioni tipo di file delimitato

È possibile specificare le seguenti impostazioni per i tipi di file delimitati.

Codifica set di caratteri

La codifica dei caratteri del file. Selezionare o specificare il nome di charset Java come "UTF-8", "ISO-8859-2", "GB18030". Il default è **UTF-8**.

Delimitatori campi

Uno o più caratteri che segnano i confini del campo. Ogni carattere viene assunto come delimitatore indipendente. Ad esempio, se si seleziona **Comma e Tab** (o si seleziona **Altro** e tipo , \t), significa che una virgola o una scheda contrassegna i limiti del campo. Se i caratteri di controllo delimitano i campi, i caratteri qui specificati vengono trattati come delimitatori oltre ai caratteri di controllo. Il default è "," se i caratteri di controllo non delimitano i campi; altrimenti il default è la stringa vuota.

I caratteri di controllo delimitano i campi

Imposta se i caratteri di controllo ASCII, ad eccezione di LF e CR, sono trattati come delimitatori di campo. Impostata per default su **No**.

Prima riga contiene nomi di campo

Imposta se utilizzare la prima riga per determinare i nomi di campo. Impostata per default su **No**.

Numero di caratteri iniziali da saltare

Il numero di caratteri all'inizio del file da saltare. Un intero non negativo. Il valore predefinito è 0.

Unire lo spazio bianco

Imposta se trattare più ricorrenze adiacenti di spazio e / o tab come un singolo delimitatore di campo. Non ha alcun effetto se né spazio né linguella è un delimitatore di campo. Il default è **Sì**.

Caratteri di commento di fine riga

Uno o più caratteri che segnano i commenti di fine riga. Il personaggio e tutto quello che lo segue sul disco vengono ignorati. Ogni carattere viene assunto come contrassegno di commento indipendente. Ad esempio, "/" * significa una barra o un asterisco avvia un commento. Non è possibile definire marcatori di commento multi - carattere come "///". La stringa vuota segnala che non sono definiti caratteri di commento. Se definiti, i caratteri di commento vengono controllati prima che i virgolette vengano elaborati o i caratteri iniziali da saltare sono saltati. Il default è la stringa vuota.

Caratteri non validi

Determina come devono essere gestiti caratteri non validi (sequenze di byte che non corrispondono ai caratteri nella codifica).

Ignora

Elimina sequenze di byte non valide.

Sostituisci con

Sostituire ogni sequenza di byte non valida con il singolo carattere.

Virgolette singole

Specifica la gestione delle virgolette singole (apostrofi). Il default è **Keep**.

Mantieni

Le virgolette singole non hanno alcun significato particolare e sono trattate come qualsiasi altro carattere.

Elimina

Le virgolette singole vengono cancellate a meno che

Coppia

Le virgolette singole sono trattate come caratteri virali e i caratteri tra coppie di singole quotazioni perdono ogni significato particolare (sono considerati citati). Se le singole virgolette stesse possono verificarsi all'interno di stringhe singole - citate è determinato dall'impostazione **Quote possono essere quotate con il raddoppio**.

Doppie virgolette

Specifica la gestione di virgolette doppie. Il default è **Pair**.

Mantieni

Le virgolette doppie non hanno alcun significato particolare e sono trattate come qualsiasi altro carattere.

Elimina

Le virgolette doppie vengono cancellate a meno che

Coppia

Le virgolette doppie sono trattate come caratteri virali e i caratteri tra coppie di virgolette doppie perdono ogni significato particolare (sono considerati citati). Se le virgolette doppie si possono verificare all'interno di stringhe doppie indicate è determinato dall'impostazione **Quote possono essere quotate raddoppiando**.

Le quotazioni possono essere citate raddoppiando

Indica se le virgolette doppie possono essere rappresentate in stringhe doppie e le singole virgolette possono essere rappresentate in stringhe singole virate quando impostata su **Pair**. Se **Sì**, le virgolette doppie sono sfuggite all'interno di stringhe doppie quotate doppie e le singole virgolette sono sfuggite all'interno di stringhe singole quotate raddoppiando. Se **No** non c'è modo di citare un doppio preventivo all'interno di una stringa a doppia quotazione o un singolo preventivo all'interno di una stringa citata da singolo. Il default è **Sì**.

Le newline possono essere sfuggite

Indica se il parser interpreta una backslash seguita dalla lettera "n", la lettera "r" o un'altra backslash come newline, ritorno a capo o carattere di backslash, rispettivamente. Se le newline non sono sfuggite, quelle sequenze di carattere vengono lette letteralmente come un backslash seguito dalla lettera "n", e così via. Il default è **No**.

Impostazioni tipo di file fisso

È possibile specificare le seguenti impostazioni per i tipi di file fissi.

Codifica set di caratteri

La codifica dei caratteri del file. Selezionare o specificare il nome di charset Java come "UTF-8", "ISO-8859-2", "GB18030". Il default è **UTF-8**.

Caratteri non validi

Determina come devono essere gestiti caratteri non validi (sequenze di byte che non corrispondono ai caratteri nella codifica).

Ignora

Elimina sequenze di byte non valide.

Sostituisci con

Sostituire ogni sequenza di byte non valida con il singolo carattere.

Lunghezza record

Indica come sono definiti i record. Se **Newline delimitato**, i record vengono definiti (delimitati) da newline, inizio del file o fine del file. Se **Lunghezza specifica**, i record sono definiti da una lunghezza record in byte. Specificare un valore positivo.

Record iniziali da saltare

Il numero di record all'inizio del file da saltare. Specificare un numero intero non negativo. Il valore predefinito è 0.

Campi

Questa sezione definisce i campi nel file. Fare clic su **Aggiungi campo** e specificare il nome del campo, la colonna in cui iniziano i valori di campo e la lunghezza dei valori di campo. Le colonne in un file sono numerate a partire da 0.

Impostazioni tipo file semi - strutturate

Le impostazioni per i file semi - strutturati sono costituite da regole per la mappatura dei contenuti del file in campi.

Tabella regole

Le singole regole estraggono informazioni da un record per creare un campo; insieme nella tabella delle regole definiscono tutti i campi che possono essere estratti da ogni record in una sorgente dati.

Le regole nella tabella vengono applicate in ordine a ogni record; se tutte le regole della tabella corrispondono al record, allora ogni altra tabella di regole non è necessaria per elaborare il record e il successivo record viene elaborato. Se una qualsiasi regola nella tabella non corrisponde, allora tutti i valori di campo estratti dalle regole precedenti nella tabella vengono scartati; se c'è un'altra tabella delle regole, le regole in quella tabella sono applicate al record. Se nessuna tabella corrisponde al record, allora viene applicata la regola di Mismatch.

Discordanza

È possibile scegliere i record **Skip** che non corrispondono a nessuna delle tabelle delle regole o impostare il valore di tutti i campi nel record a **Manca** (null).

Esporta regole

È possibile salvare la tabella delle regole attualmente visibili per il riutilizzo. La tabella esportata viene salvata sul server.

Importa regole

È possibile importare una tabella delle regole salvate nella tabella delle regole attualmente visibili. Questo sovrascrive qualsiasi regola definita per quella tabella, quindi è meglio creare una nuova tabella e quindi importare una tabella di regole.

Editor delle regole

L'editor di regole consente di creare una regola di estrazione per un singolo campo.

Gruppo di acquisizione anonimo

Una regola di acquisizione del campo tipicamente inizia ad estrarre i dati da un record nella posizione in cui la regola precedente si è fermata. Quando ci sono informazioni estranee tra due campi in una fonte di dati semi - strutturata, può quindi essere utile definire un gruppo di acquisizione anonimo che posiziona il parser dove inizia il campo successivo. Quando si seleziona **Anonymous capture group**, i controlli per la denominazione e l'etichettatura del gruppo di cattura sono disabilitati, ma il resto delle funzioni di dialogo normalmente.

Nome campo

Immettere un nome per il campo. Questo viene utilizzato per definire i metadati dell'origine dati. I nomi di campo devono essere univoci all'interno di una tabella regole.

Nome regola

Facoltativamente inserire un'etichetta descrittiva per la regola.

Descrizione

Opzionalmente inserire una descrizione più lunga per la regola.

Definizione di una regola

Ci sono due metodi per definire le regole.

Controlli di utilizzo per le regole di estrazione

Ciò semplifica la creazione di regole di estrazione.

1. Specificare il punto per iniziare l'estrazione dei dati sul campo; **Posizione corrente** inizierà dove la regola precedente si è arrestata, e **Skip fino a** partirà all'inizio del record e ignorerà tutti i caratteri fino a quando non raggiungerà quello specificato nella casella di testo. Selezionare **Includi** se si desidera che i dati sul campo includa il carattere in posizione di inizio.
2. Selezionare un gruppo di acquisizione del campo dal dropdown **Capture**.
3. Opzionalmente selezionare il punto per interrompere l'estrazione dei dati sul campo; **Whitespace** si fermerà quando si incontreranno eventuali caratteri di whitespace (come gli spazi o le schede) e **A carattere (s)** si fermerà alla stringa specificata. Selezionare **Includi** se si desidera che i dati sul campo includa il carattere alla posizione di arresto.

Definire manualmente le regole regexp

Selezionare questo se si è a proprio agio a scrivere una sintassi di espressione regolare. Inserire un'espressione regolare nella casella di testo **Regexp**.

Aggiungi gruppo Capture Field

Ciò consente di salvare l'espressione regolare per un utilizzo successivo. Il gruppo di cattura salvato appare sul dropdown **Capture**.

L'Editor delle regole mostra un'anteprima dei dati estratti dal primo record da questa regola, dopo che sono state applicate tutte le regole precedenti nella tabella delle regole.

Impostazioni tipo file excel

È possibile specificare le seguenti impostazioni per i file Excel.

Selezione del foglio di lavoro

Seleziona il foglio di lavoro Excel da utilizzare come fonte dati. Specificare un indice numerico (l'indice del primo foglio di lavoro è 0) o il nome del foglio di lavoro. Il default è quello di utilizzare il primo foglio di lavoro.

Selezione intervallo dati per l'importazione.

È possibile importare i dati iniziando dalla prima riga non vuota o da un intervallo esplicito di celle.

- **L'intervallo inizia alla prima riga non vuota.** Individua la prima cella non vuota e la utilizza come angolo superiore sinistro dell'intervallo di dati.
- In alternativa specificare una gamma esplicita di celle per riga e colonna. Ad esempio, per specificare l'intervallo Excel A1:D5, è possibile inserire A1 nel primo campo e D5 nella seconda (o in alternativa, R1C1 e R5C4). Vengono restituite tutte le righe dell'intervallo specificato, incluse quelle vuote.

Prima riga contiene nomi di campo

Specifica se la prima riga della gamma di celle selezionate contiene i nomi di campo. Il default è **No**.

Stop alla lettura dopo aver incontrato righe vuote

Specifica se arrestare i record di lettura dopo che si incontra più di una riga vuota o continuare a leggere tutti i dati alla fine del foglio di lavoro, incluse le righe vuote. Il default è **No**.

Formati

La Scheda Formati consente di definire le informazioni di formattazione per i campi paralizzanti.

Impostazioni Conversione Field

Spazio bianco Trim

Rimuove i caratteri di spazio bianco dall'inizio e / o dalla fine dei campi di stringa. Default su **Nessuno**. Sono supportati i seguenti valori:

Nessuna

Non elimina i caratteri dello spazio bianco.

A sinistra

Rimuove i caratteri dello spazio bianco dall'inizio della stringa.

Destra

Rimuove i caratteri di spazio bianco dalla fine della stringa.

Entrambi

Rimuove i caratteri dello spazio bianco dall'inizio e dalla fine della stringa.

Impostazioni internazionali

Definisce un locale. Default sul locale del server. La stringa locale deve essere specificata come: < language> [_paese [_variante]], dove:

lingua

Un codice valido, inferiore, a due lettere come definito da ISO-639.

nazione

Un codice valido, in maiuscolo, a due lettere come definito da ISO-3166.

variante

Un codice specifico del fornitore o del browser.

Separatore decimale

Imposta il carattere utilizzato come segno decimale. Impostata per default sull'impostazione specifica del locale.

Raggruppamento dei simboli

Imposta se deve essere utilizzato o meno il carattere locale - specifico utilizzato per il separatore di migliaia.

Formato data predefinito

Definisce un formato data predefinito. Sono supportati tutti i modelli di formato definiti dalla specifica [unicode data markup language \(LDML\)](#) .

Formato orario predefinito

Definisce un formato di tempo predefinito.

Data orario predefinita

Definisce un formato di data / ora predefinito.

Fuso orario predefinito

Imposta il fuso orario. Default su UTC. L'impostazione si applica ai campi del tempo e della registrazione data / ora che non hanno esplicitamente specificato timezone.

Sostituzioni campo

Questa sezione consente di assegnare le istruzioni di formattazione ai singoli campi. Selezionare un campo dal modello dati, oppure digitare un nome di campo e fare clic su **Aggiungi** per aggiungerlo all'elenco dei campi con istruzioni individuali. Fare clic su **Rimuovi** per rimuoverlo dall'elenco. Per un campo selezionato nell'elenco è possibile impostare le seguenti proprietà del campo.

Archiviazione

Impostare la memoria del campo.

Separatore decimale

Per i campi con memoria Reale, imposta il carattere utilizzato come segno decimale. Impostata per default sull'impostazione specifica del locale.

Raggruppamento dei simboli

Per i campi con Integer o Real storage è necessario utilizzare o meno il carattere locale - specifico utilizzato per il separatore di migliaia.

Formati

Per i campi con data, ora o archiviazione Data / ora, imposta il formato. Scegli un formato dall'elenco a discesa.

Scheda Ordine campo

Per i tipi di file delimitati e Excel, la scheda Ordine dei campi consente di definire l'ordine paralizzato dei campi per il file. Questo è importante quando ci sono più file in un'origine dati, perché l'ordine effettivo dei campi può essere diverso attraverso i file, ma l'ordine di parati dei campi deve essere lo stesso per creare un modello di dati coerente. Per i tipi di file fissi e semi - strutturati, l'ordine viene definito nella scheda Impostazioni.

Quando c'è un singolo file nell'origine dati o tutti i file hanno lo stesso ordine di campo, è possibile utilizzare l'impostazione predefinita **Field order corrisponde al modello di dati**. Se ci sono più file nell'origine dati e l'ordine dei campi nel file non corrisponde, definire un **ordine di campo specifico** per l'analisi del file.

1. Per aggiungere un campo all'elenco ordinato, digitare il nome del campo o selezionarlo dall'elenco fornito dal modello dati. È possibile aggiungere tutti i campi nel modello dati contemporaneamente facendo clic su **Aggiungi tutto**. I nomi di campo verranno aggiunti solo una volta all'elenco ordinato.
2. Utilizzare i pulsanti freccia per ordinare i campi come desiderato.

Quando viene utilizzato **Ordinamento campo specifico**, eventuali campi non aggiunti all'elenco non fanno parte della serie di risultati per questo file. Se ci sono campi nel modello dati che non sono elencati in questa finestra di dialogo, i valori sono nulli nella serie di risultati.

Scheda cartella

Quando si specificano le impostazioni del parser per una cartella, la scheda Cartella consente di scegliere quali file nella cartella sono inclusi nell'origine dati.

Corrisponde a tutti i file nella cartella selezionata

L'origine dati include tutti i file nel livello superiore della cartella; i file nelle sottocartelle non sono inclusi.

Abbinare i file utilizzando un'espressione regolare

L'origine dati include tutti i file nel livello superiore della cartella che corrispondono all'espressione regolare specificata; i file nelle sottocartelle non sono inclusi.

Corrisponde ai file che utilizzano un'espressione di globbing Unix (potenzialmente ricorsiva)

L'origine dati include tutti i file che corrispondono all'espressione di globbing Unix specificata; l'espressione può includere file che si trovano nelle sottocartelle della cartella selezionata.

Associazioni Campo HCatalog

Schema HCatalog

Visualizza la struttura della tabella specificata. HCatalog può supportare un dataset altamente strutturato. Per definire una fonte dati Server analitico su tali dati, la struttura deve essere appiattita in righe e colonne semplici. Selezionare un elemento nello schema e fare clic sul pulsante di spostamento per associarlo ad un campo per l'analisi.

Non tutti i nodi dell'albero possono essere mappati. Ad esempio, una schiera o mappa di tipi complessi è considerata un "genitore" e non può essere direttamente mappata; ogni elemento semplice in una schiera o mappa di HCatalog deve essere aggiunto separatamente. Questi nodi possono essere identificati dall'etichetta nell'albero che termina in ... :array:struct o ... :map:struct.

Ad esempio:

- Per una schiera di interi, è possibile assegnare un campo a un valore all'interno della matrice: bigintarray [45], ma non la matrice stessa: bigintarray

- Per una mappa è possibile assegnare un campo a un valore all'interno della mappa: datamap ["key"], ma non la mappa stessa: datamap
- Per una schiera di una schiera di interi, è possibile assegnare un campo a un bigintarrayarray di valore [45] [2], ma non la matrice stessa, bigintarrayarray [45].

Pertanto, quando si assegna un campo ad un elemento schiera o mappa, la definizione dell'elemento deve includere l'indice o la chiave: bigintarray [index] o bigintmap ["key"].

L'utente corrente può vedere solo le tabelle a cui ha accesso. Poiché la directory HDFS è l'unica directory con autorizzazione di lettura ed esecuzione (il file interno ha il permesso di lettura, che l'utente può vedere), gli utenti non possono vedere tabelle a cui non hanno accesso. La limitazione è in atto per proteggere le tabelle di alveare gestite, le tabelle di alveare esterne e le directory partizionate.

Associazioni campi

Elemento di HCatalog

Fare doppio clic su una cella da modificare. È necessario modificare la cella quando l'elemento HCatalog è una matrice o una mappa. Con un array, specificare il numero intero corrispondente al membro della schiera che si desidera mappare in un campo. Con una mappa, specificare una stringa citata corrispondente alla chiave che si desidera mappare in un campo.

Campo di associazione

Il campo come appare nella sorgente dati Server analitico . Fare doppio clic su una cella da modificare. I valori duplicati nella colonna Mapping Field non sono ammessi e risultano in un errore.

Archiviazione

Lo stoccaggio del campo. La memoria deriva da HCatalog e non può essere modificata.

Nota: Quando si fa clic su [Anteprima e Metadati](#) per finalizzare un'origine dati HCatalog non ci sono opzioni di modifica.

Dati non elaborati

Visualizza i record in quanto sono memorizzati in HCatalog; questo può aiutare a determinare come associare lo schema HCatalog ai campi.

Nota: Qualsiasi filtraggio specificato nelle Selezioni HCatalog viene applicato alla vista dei dati greli.

Utilizzo delle origini dati HCatalog

Server analitico fornisce il supporto per le origini dati HCatalog. Questa sezione descrive come configurare diversi database NoSQL sottostanti.

Nella maggior parte dei casi, è necessario fare riferimento alla documentazione del fornitore per l'integrazione Hive .

Apache Accumulo

<https://cwiki.apache.org/confluence/display/Hive/AccumuloIntegration>

Apache Cassandra

[“Apache Cassandra” a pagina 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

Origini dati XML

[“Origini dati XML” a pagina 15](#)

Apache Cassandra

Server analitico fornisce supporto per le origini dati HCatalog che hanno contenuto sottostante in Apache Cassandra. Cassandra fornisce un archivio chiavi - value strutturato. Chiavi mappa a più valori, che sono raggruppate in famiglie di colonne. Le famiglie di colonne sono fissate quando viene creato un database, ma le colonne possono essere aggiunte a una famiglia in qualsiasi momento. Inoltre, le colonne vengono aggiunte solo a chiavi specificate, quindi le diverse chiavi possono avere diversi numeri di colonne in qualsiasi famiglia data. I valori di una famiglia di colonne per ogni chiave vengono memorizzati insieme.

Ci sono due modi per definire le tabelle Cassandra : utilizzando l'interfaccia riga di comando legacy Cassandra (cassandra-cli) e la nuova shell CQL (cqlsh).

Utilizzare la seguente sintassi per creare una tabella esterna Apache Cassandra in Hive se la tabella è stata creata utilizzando CLI legacy.

```
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>");
```

Ad esempio per la seguente definizione di tabella CLI:

```
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'];
```

... il DDL da tavolo Hive sarà così:

```
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");
```

Utilizzare la seguente sintassi per creare una tabella esterna Apache Cassandra in Hive se la tabella è stata creata utilizzando CQL.

```
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>");
```

Ad esempio per la seguente definizione di tabella CQL3 :

```
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);

```

... il DDL da tavolo Hive è il seguente:

```

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");

```

Origini dati XML

Server analitico fornisce supporto per i dati XML tramite HCatalog.

Esempio

1. Associare lo schema XML ai tipi di dati Hive attraverso il DDL (Data Definition Language) Hive , secondo le seguenti regole.

```

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.q1.io.IgnoreKeyTextOutputFormat"
[LOCATION "<data_location>"]
TBLPROPERTIES (
  "xmlinput.start"="<start_tag " ,
  "xmlinput.end"="<end_tag>"
);

```

Nota: Se i tuoi file XML sono compressi con la compressione Bz2 , quindi INPUTFORMAT dovrebbe essere impostato su `com.ibm.spss.hive.serde2.xml.SplittableXmlInputFormat`. Se vengono compressi con compressione CMX, dovrebbe essere impostato su `com.ibm.spss.hive.serde2.xml.CmxXmlInputFormat`.

Ad esempio, il seguente XML ...

```
<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>
```

... sarebbe rappresentato dal seguente Hive DDL.

```
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>"
);
```

Per ulteriori informazioni, consultare [“XML a Hive Data Tipi Mapping”](#) a pagina 16.

2. Creare un'origine dati Server analitico con tipo di contenuto HCatalog nella Server analitico Console.

Limitazioni

- Solo la specifica XPath 1.0 è attualmente supportata.
- La parte locale dei nomi qualificati per gli elementi e gli attributi vengono utilizzati quando si gestiscono i nomi di campo Hive . I prefissi dello spazio dei nomi vengono ignorati.

XML a Hive Data Tipi Mapping

I dati modellati in XML possono essere trasformati nei tipi di dati Hive utilizzando le convenzioni riportate di seguito.

Strutture

L'elemento XML può essere direttamente mappato sul tipo di struttura Hive in modo che tutti gli attributi diventino membri dei dati. Il contenuto dell'elemento diventa un membro aggiuntivo di tipo primitivo o complesso.

dati XML

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

Hive DDL e dati raw

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

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

Array

Le sequenze XML di elementi possono essere rappresentate come array Hive di tipo primitivo o complesso. Il seguente esempio mostra come l'utente può definire una schiera di stringhe utilizzando il contenuto dell'elemento XML `<result>`.

dati XML

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

Hive DDL e dati raw

```
result array<string>
```

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

Mappe

Lo schema XML non fornisce supporto nativo per le mappe. Ci sono tre approcci comuni per modellare le mappe in XML. Per alloggiare i diversi approcci utilizziamo la seguente sintassi:

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

dove

element_name

Il nome dell'elemento XML da considerare come una voce mappa

chiave

Il nodo XML key entry key

valore

Il nodo XML del valore entry value

La specifica di mappa per l'elemento XML dato deve essere definita nella sezione SERDEPROPERTIES nella DDL di creazione della tabella Hive. I tasti e i valori possono essere definiti utilizzando la seguente sintassi:

@attribute

La specifica @attribute consente all'utente di utilizzare il valore dell'attributo come chiave o valore della mappa.

elemento

Il nome elemento può essere utilizzato come chiave o valore.

#content

Il contenuto dell'elemento può essere utilizzato come chiave o valore. Poiché le chiavi di mappa possono essere solo di tipo primitivo il contenuto complesso verrà convertito in stringa.

Gli approcci per rappresentare le mappe in XML, e il loro corrispondente Hive DDL e i dati greli, sono i seguenti.

Nome elemento ai contenuti

Il nome dell'elemento viene utilizzato come chiave e il contenuto come valore. Questa è una delle tecniche comuni e viene utilizzata per impostazione predefinita quando si mappano XML ai tipi di mappa Hive. L'evidente limitazione con questo approccio è che la chiave di mappa può essere solo di tipo stringa.

dati XML

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

Mappatura, Hive DDL e dati raw

In questo caso non è necessario specificare una mappatura perché il nome dell'elemento viene utilizzato come chiave e il contenuto come valore per impostazione predefinita.

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

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

Attributo a Content Content

Utilizzare un valore di attributo come chiave e il contenuto elemento come valore.

dati XML

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

Mappatura, Hive DDL e dati raw

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

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

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

Attributo all'attributo

dati XML

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

Mappatura, Hive DDL e dati raw

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

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

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

Contenuto complesso

Il contenuto complesso utilizzato come tipo primitivo verrà convertito in una stringa XML valida aggiungendo un elemento root chiamato < string>. Considerare il seguente XML:

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

L'espressione XPath /dataset/ * comporterà un numero di < value> nodi XML restituiti. Se il campo di destinazione è di tipo primitivo l'implementazione trasformerà il risultato della query all'XML valido aggiungendo il nodo root < string>.

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

Nota: L'implementazione non aggiungerà un elemento root < string> Se il risultato della query è un singolo elemento XML.

Contenuto testo

Il whitespace - solo il contenuto di testo di un elemento XML viene ignorato.

Anteprima e Metadati (origini dati)

Cliccando su **Anteprima e Metadati** viene visualizzato un esempio di record e il modello di dati per l'origine dati. Qui hai la possibilità di rivedere le informazioni di base dei metadati.

Anteprima

La scheda Anteprima mostra un piccolo campione di record e i relativi valori di campo.

Modifica

La scheda Modifica visualizza i metadati del campo base. Per le origini dati con il tipo di contenuto Files, il modello di dati viene generato da un piccolo campione di record ed è possibile modificare manualmente i metadati del campo su questa scheda. Per le origini dati con tipo di contenuto HCatalog, il modello di dati viene generato in base alle Mappature di campo HCatalog e non è possibile modificare la memoria del campo su questa scheda.

Campo

Fare doppio clic sul nome del campo per modificarlo.

Misurazione

Livello di misurazione utilizzato per descrivere le caratteristiche dei dati in un campo specifico.

Ruolo

Utilizzato per raccontare i nodi di modellazione se i campi saranno Input (campi predittivi) o Target (campi previsti) per un processo di machine - learning. Entrambi e Nessuno sono altri ruoli disponibili, oltre a Partizione, che indica un campo utilizzato per eseguire la partizione dei record in campioni separati per addestramento, test e convalida. Il valore Split specifica che verranno costruiti modelli separati per ogni possibile valore del campo. La frequenza specifica che un valore di campo deve essere utilizzato come peso di frequenza per ogni record. L'ID record viene utilizzato per identificare un record nell'output.

Archiviazione

L'archiviazione definisce le modalità di archiviazione dei dati in un campo. Per esempio, un campo con valori 1 e 0 è un campo in cui sono archiviati dati di tipo intero. Si distingue dal livello di misurazione, che descrive l'utilizzo dei dati e non influisce sull'archiviazione. Per esempio, è possibile impostare il livello di misurazione per un campo intero con valori 1 e 0 su Indicatore. In genere, questo indica che 1=Vero e 0=Falso.

Valori

Mostra i valori individuali per i campi con misurazione categoriale o la gamma di valori per i campi con misurazione continua.

Struttura

Indica se i record nel campo contengono un valore unico (Primitivo) o un elenco di valori.

Profondità

Indica la profondità di un elenco; 0 è un elenco di primitivi, 1 è un elenco di liste e così via.

Scansione tutti i valori dei dati

Ciò consente di avviare ed annullare la scansione dei valori dei dati di origine dati per determinare i valori di categoria e i limiti di gamma. Se una scansione è in corso, fare clic sul pulsante su **Annulla scansione dati**. La scansione di tutti i valori dei dati garantisce che i metadati siano corretti, ma può richiedere un po' di tempo se l'origine dati ha molti campi e record.

Progetti

I progetti sono spazi di lavoro per la memorizzazione di input e l'accesso all'output dei lavori. Forniscono la struttura organizzativa di primo livello per contenere file e cartelle. I progetti possono essere condivisi con singoli utenti e gruppi.

Elenco dei progetti

La pagina Progetti principali fornisce un elenco di progetti di cui l'utente corrente è membro.

- Clicca sul nome di un progetto per visualizzarne i dettagli e modificarne le proprietà.
- Digitare nell'area di ricerca per filtrare l'elenco per visualizzare solo i progetti con la stringa di ricerca nel loro nome.
- Clicca su **Nuovo** per creare un nuovo progetto con il nome specificato nella finestra di dialogo **Aggiungi nuovo progetto**. Vedi [“Regole di denominazione” a pagina 23](#) per le limitazioni ai nomi che puoi dare ai progetti.
- Clicca su **Elimina** per rimuovere il progetto selezionato. Questa azione rimuove il progetto e cancella tutti i dati associati al progetto da HDFS.
- Clicca su **Aggiorna** per aggiornare la quotazione.

Dettagli del progetto individuale

L'area dei contenuti è suddivisa in sezioni comprimibili **Dettagli**, **Sharing**, **File** e **Versioni**.

Dettagli

Nome

Un campo di testo modificabile che mostra il nome del progetto.

Nome visualizzato

Un campo di testo modificabile che mostra il nome del progetto come visualizzato in altre applicazioni. Se questo è vuoto, il Nome viene utilizzato come nome di visualizzazione.

Descrizione

Un campo di testo modificabile per fornire un testo esplicativo sul progetto.

Versioni da conservare

Elimina automaticamente la versione del progetto commessa più grande quando il numero delle versioni supera il numero specificato. Il valore predefinito è 25.

Nota: Il processo di pulizia non è immediato, ma corre sullo sfondo ogni 20 minutes minuti.

È pubblico

Una casella di controllo che indica se qualcuno può vedere il progetto (controllato) o se gli utenti e i gruppi devono essere esplicitamente aggiunti come membri (sdoganati).

Clicca su **Salva** per mantenere lo stato attuale delle impostazioni.

Condivisione

È possibile condividere un progetto aggiungendo utenti e gruppi come autori o telespettatori.

- Digitando nella casella di testo filtri su utenti e gruppi con la stringa di ricerca nel loro nome. Selezionare il livello di condivisione e fare clic su **Aggiungi membro** per aggiungere all'elenco dei membri.
 - Gli autori sono membri completi di un progetto, e possono modificare il progetto così come le cartelle e i file al suo interno. Questi utenti e membri di questi gruppi hanno accesso alla scrittura (Server analitico Export node) a questo progetto quando si connette a Server analitico tramite IBM SPSS Modeler.
 - I telespettatori possono vedere le cartelle e i file all'interno di un progetto, e definire le fonti di dati sugli oggetti all'interno di un progetto, ma non possono modificare il progetto.
- Per rimuovere un autore, selezionare un utente o un gruppo nell'elenco Autore e fare clic su **Rimuovi membro**.

Nota: Gli amministratori hanno letto e scrivono l'accesso ad ogni progetto, indipendentemente dal fatto che siano specificamente elencati come membro.

Nota: Le modifiche apportate a Sharing vengono applicate immediatamente e automaticamente.

File

Riquadro struttura progetto

Il riquadro a destra mostra la struttura progetto / cartella per il progetto attualmente selezionato. È possibile sfogliare la struttura della cartella, ma non è modificabile, tranne che attraverso i pulsanti.

- Fare clic su **Scarica file sul file system locale** per scaricare un file selezionato sul file system locale.
- Clicca su **Elimina i file selezionati** per rimuovere il file / cartella selezionato.

File Viewer

Mostra la struttura della cartella per il progetto corrente. La struttura della cartella è solo modificabile all'interno di progetti definiti. Ovvero non è possibile aggiungere file, creare cartelle o cancellare elementi a livello di root della modalità **Progetti**. Per creare la cancellazione di un progetto, tornare alla quotazione del Progetto.

- Clicca su **Carica file su HDFS** per caricare un file nella cartella / sottocartella corrente.
- Fare clic su **Crea una nuova cartella** per creare una nuova cartella sotto la cartella corrente, con il nome specificato nella finestra di dialogo **Nuovo nome cartella**.
- Clicca su **Scarica file sul file system locale** per scaricare i file selezionati sul file system locale.
- Clicca su **Elimina i file selezionati** per rimuovere i file / cartelle selezionati.

Versioni

I progetti sono versati in base alle modifiche al contenuto del file e della cartella. Le modifiche agli attributi di un progetto, come la descrizione, che sia pubblica, e con la quale è condivisa, non richiedono una nuova versione. L'aggiunta, la modifica o la cancellazione di file o cartelle richiede una nuova versione.

Tabella di versioning del progetto

La tabella visualizza le versioni di progetto esistenti, le relative date di creazione e di impegno, gli utenti responsabili di ogni versione e la versione principale. La versione principale è la versione su cui si basa la versione selezionata.

- Clicca su **Blocca** per apportare modifiche al contenuto della versione del progetto selezionato.
- Clicca su **Commit** per salvare tutte le modifiche apportate ad un progetto e rendere questa versione lo stato attuale visibile del progetto.
- Clicca su **Discard** per scartare tutte le modifiche apportate ad un progetto bloccato e restituire lo stato visibile del progetto alla versione più recente impegnata.
- Clicca su **Elimina** per rimuovere la versione selezionata.

Gestione degli utenti e dei gruppi

Gli amministratori possono creare, eliminare e, modificare utenti e gruppi attraverso il portale Server analitico User Management.

1. Dopo la registrazione nella console Server analitico, espandere il menu a tendina accanto al tuo ID di login (posizionato vicino al top - destra) e selezionare **Gestione utenti**. Il portale Server analitico User Management si apre in una finestra del browser separato.
2. Inserire le proprie credenziali Server analitico User Management e fare clic su **Login**.

Server analitico Opzioni di gestione utenti

Il portale è diviso tra le sezioni comprimibili **Gestisci utente** e **Gestisci gruppo**.

Gestisci utente

Le opzioni includono l'aggiunta di nuovi utenti, la modifica degli utenti esistenti, l'eliminazione degli utenti esistenti e la visualizzazione dell'elenco degli utenti primari.

Aggiungi utente

Nome utente

Inserire un nome utente Server analitico valido.

Password

Inserire una password per il nome utente specificato.

Conferma password

Immettere nuovamente la password.

Gruppi

Facoltativamente, selezionare un gruppo esistente dall'elenco a discesa. Dopo aver fatto clic su **Invia**, l'utente diventerà membro del gruppo selezionato.

Elenco utenti

Visualizza tutti gli utenti in una tabella, elencati per nome utente e associazione di gruppo.

- Utilizzare il campo **Ricerca** per cercare i nomi degli utenti all'interno della tabella.
- Utilizzare i controlli radio selezionati più utenti. Clicca su **Elimina** per rimuovere gli utenti selezionati.
- Modifica le proprietà degli utenti cliccando su un nome utente.

Nota: Il valore **Nome utente** non può essere modificato.

Gestisci gruppo

Le opzioni includono l'aggiunta di nuovi gruppi, la modifica dei gruppi esistenti, l'eliminazione dei gruppi esistenti e la visualizzazione dell'elenco dei gruppi primari.

Aggiungi gruppo

Nome gruppo

Inserire un nome di gruppo.

Utenti

Selezionare gli utenti da aggiungere al gruppo dall'elenco a discesa. Clicca su **Invia** per salvare le impostazioni del gruppo.

Elenco gruppi

Visualizza tutti i gruppi in una tabella, elencati per nome di gruppo e utenti all'interno di ciascun gruppo.

- Utilizzare il campo **Ricerca** per cercare i nomi di gruppo all'interno della tabella.
- Utilizzare i controlli radio selezionati più gruppi. Clicca su **Elimina** per rimuovere i gruppi selezionati.
- Modificare le proprietà del gruppo facendo clic su un nome di gruppo.

Nota: Il valore **Nome gruppo** non può essere modificato.

Gestione dei ruoli utente e di gruppo

Gli amministratori possono gestire i ruoli di utenti e gruppi attraverso la pagina Utenti.

L'area dei contenuti è suddivisa in sezioni comprimibili **Dettagli** e **Principali**.

Dettagli

Nome

Un campo di testo non modificabile che visualizza il nome dell'inquilino.

Descrizione

Un campo di testo modificabile che consente di fornire un testo esplicativo sull'inquilino.

URL (Uniform Resource Locator)

Questo è l'URL da dare agli utenti per accedere all'inquilino tramite la console Server analitico .

Stato

Gli inquilini **Active** sono attualmente in uso. Fare un tenant **Inattivo** impedisce agli utenti di accedere a quell' inquilino, ma non elimina nessuna delle informazioni sottostanti.

Principali

I principal sono utenti e gruppi che vengono disegnati dal provider di sicurezza che viene impostato durante la configurazione. È possibile modificare il ruolo dei principal per essere Amministratori, Utenti o Readers.

Metriche

Consente di configurare i limiti delle risorse per un inquilino. Riporta lo spazio su disco attualmente utilizzato dall'inquilino.

- È possibile impostare una quota massima di spazio su disco per l'inquilino; quando questo limite ha raggiunto, non è possibile scrivere più dati su disco su questo tenant fino a quando non viene sdoganato abbastanza spazio su disco per portare l'utilizzo dello spazio del disco tenant sotto la quota.
- È possibile impostare un livello di avvertenza dello spazio su disco per l'inquilino; quando la quota viene superata, nessun lavoro analitico può essere inoltrato dai principal su questo tenant fino a quando non viene sdoganato abbastanza spazio su disco per portare l'utilizzo dello spazio del disco tenant sotto la quota.
- È possibile impostare un numero massimo di lavori paralleli che possono essere eseguiti in un'unica volta su questo tenant; quando la quota viene superata, nessun lavoro analitico può essere inoltrato dai principal su questo tenant fino a quando non si completa un lavoro attualmente in esecuzione.
- È possibile impostare il numero massimo di campi che un'origine dati può avere. Il limite viene controllato ogni qualvolta viene creata o aggiornata un'origine dati.
- È possibile impostare il numero massimo di record che una fonte di dati può avere. Il limite viene controllato ogni qualvolta viene creata o aggiornata un'origine dati; ad esempio, quando si aggiungono un nuovo file o le impostazioni di modifica per un file.
- È possibile impostare la dimensione massima del file in megabyte. Il limite viene controllato quando viene caricato un file.

Configurazione provider di sicurezza

Consente di specificare il provider di autenticazione utente. **Default** utilizza il provider di tenant predefinito, impostato durante l'installazione e la configurazione. **LDAP** consente di autenticare gli utenti con un server LDAP esterno, ad esempio Active Directory o OpenLDAP. Specificare le impostazioni per il provider e facoltativamente specificare le impostazioni del filtro per controllare gli utenti e i gruppi disponibili nella sezione Principali.

Regole di denominazione

Per qualsiasi cosa possa essere dato un nome univoco in Server analitico, come ad esempio le origini dati e i progetti, a quei nomi vengono applicate le seguenti regole.

- All'interno di un unico inquilino, i nomi devono essere unici all'interno di oggetti dello stesso tipo. Ad esempio, due origini dati non possono essere entrambe denominate insuranceClaims, ma un'origine dati e un progetto possono essere denominati insuranceClaims.
- I nomi rispettano la distinzione maiuscole/minuscole. Ad esempio, insuranceClaims e InsuranceClaims sono considerati nomi univoci.
- I nomi ignorano lo spazio bianco leader e trailing.
- I seguenti caratteri non sono validi nei nomi.

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

Capitolo 2. SPSS Modeler Integrazione

SPSS Modeler è un workbench di data mining che ha un approccio visivo all'analisi. Ogni azione distinta in un lavoro, dall'accesso ad una fonte di dati per unire record per scrivere un nuovo file o costruire un modello, è rappresentata da un nodo sulla canvas. Collettiamo queste azioni insieme per formare un flusso analitico. Per costruire un flusso analitico che corre con Server analitico:

1. Il flusso deve iniziare con un nodo Server analitico Source.
2. Costruire la metà del flusso nell'interfaccia Modeler come si farebbe normalmente, scegliendo nodi di processo (Field o Record Ops) supportati da Server analitico. Esiste un pannello Server analitico nella tavolozza Modeler che visualizza i nodi supportati.
3. Ci sono un paio di opzioni per finire il flusso.
 - Scegliere un nodo terminale (Output, Graph, Export o Modeling) supportato da Server analitico. In questo caso, Modeler spinge l'intero flusso su Server analitico. Server analitico orchestra i lavori necessari sul cluster Hadoop e rende disponibili i risultati a Modeler. Modeler prende i risultati e li presenta a te, proprio come farebbe se il flusso venisse elaborato localmente.
 - Se si sceglie un nodo terminale non supportato da Server analitico, Modeler spinge quanto più possibile il flusso verso Server analitico, quindi inizia a estrarre record da Hadoop. Si noti che i modelli che attualmente non possono essere costruiti con Server analitico possono essere segnati da Server analitico. Questo significa che puoi strutturare un flusso per prendere un sottocampione statisticamente valido dei tuoi big data con Server analitico, quindi costruire un modello "localmente" in Modeler. Il nugget modello risultante può quindi essere incluso in un flusso di punteggio che viene eseguito interamente in Server analitico.

Nota: È possibile impostare il numero massimo di record SPSS Modeler che verrà scaricato da Hadoop nelle proprietà stream Server analitico.

Nodi supportati

Molti nodi SPSS Modeler sono supportati per l'esecuzione su HDFS, ma potrebbero esserci delle differenze nell'esecuzione di determinati nodi ed alcuni non sono attualmente supportati. Questa sezione illustra i livelli di supporto attuali.

Nota: Consultare la documentazione di SPSS Modeler per informazioni sul regolare funzionamento di questi nodi.

Generale

- Alcuni caratteri, che normalmente sono accettabili all'interno di un nome campo Modeler racchiuso tra virgolette, non verranno accettati da Server analitico.
- Per poter essere eseguito in Server analitico, un flusso Modeler deve iniziare con uno o più nodi origine Server analitico e terminare in un singolo nodo modellazione o nodo esportazione di Server analitico.
- Si consiglia di impostare l'archiviazione dei target continui come reale piuttosto che intero. I modelli di calcolo del punteggio scrivono sempre valori reali nei file di dati di output per i target continui, mentre il modello di dati di output per i punteggi segue l'archiviazione del target. Quindi, se un target continuo ha un'archiviazione numero intero, si verificherà una mancata corrispondenza nei valori scritti e il modello dati per i punteggi e tale mancata corrispondenza causerà errori durante il tentativo di leggere i dati totalizzati.

Origine

- Un flusso che inizia con un nodo origine diverso da Server analitico verrà eseguito a livello locale.

Operazioni su record

Sono supportate tutte le operazioni sui record, ad eccezione dei nodi Streaming TS e STB (Space-Time-Boxes). Di seguito sono riportate ulteriori note sulla funzionalità dei nodi supportati.

Selezione

- Supporta lo stesso insieme di funzioni supportato dal [nodo Ricava](#).

Esempio

- Il campionamento a livello di blocco non è supportato.
- I metodi di campionamento complessi non sono supportati.
- Il primo campionamento n con "Discard sample" non è supportato.
- Il primo campionamento n con $N > 20000$ non è supportato.
- Il campionamento da 1 in n non è supportato quando "Maximum sample size" non è impostato.
- Il campionamento da 1 in n non è supportato quando $N * \text{"Maximum sample size"} > 20000$.
- Il campionamento a livello di blocco % casuale non è supportato.
- La % casuale attualmente supporta la fornitura di un seed.

Aggregato

- Le chiavi contigue non sono supportate. Se si sta riutilizzando un flusso esistente configurato per ordinare i dati e quindi si utilizza questa impostazione nel nodo Aggregato, cambiare il flusso per rimuovere il nodo Ordina.
- Le statistiche dell'ordine (mediana, 1° quartile e 3° quartile) sono calcolate in modo approssimativo e supportate mediante la scheda Ottimizzazione.

Ordinamento

- La scheda Ottimizzazione non è supportata.

In un ambiente distribuito, vi è un numero limitato di operazioni che mantiene l'ordine di record stabilito dal nodo Ordina.

- Un nodo Ordina seguito da un nodo di esportazione produce un'origine dati ordinata.
- Un nodo Ordina seguito da un nodo Campione con campionamento del **Primo** record, restituisce i primi *N* record.

In generale, collocare un nodo Ordina il più vicino possibile alle operazioni che necessitano dei record ordinati.

Unione

- L'unione in base all'ordine non è supportata.
- La scheda Ottimizzazione non è supportata.
- Le operazioni di unione sono relativamente lente. Se si dispone di spazio sufficiente in HDFS, può essere molto più rapido unire le origini dati una volta sola ed utilizzare l'origine unita nei flussi successivi che unire le origini dati in ogni flusso.

Trasformazione R

La sintassi R nel nodo deve essere costituita da operazioni record-at-a-time.

Operazioni su campi

Sono supportate tutte le operazioni sui campi, ad eccezione dei nodi Anonimizza, Trasponi, Intervalli di tempo e Cronologia. Di seguito sono riportate ulteriori note sulla funzionalità dei nodi supportati.

Preparazione automatica dei dati

- L'addestramento del nodo non è supportato. L'applicazione delle trasformazioni in un nodo Preparazione automatica dei dati addestrato a nuovi dati è supportata.

Ricava

- Tutte le funzioni di Ricava sono supportate, ad eccezione delle funzioni di sequenza.
- La derivazione di un nuovo campo come Conteggio è essenzialmente un'operazione di sequenza, e quindi non supportata.
- I campi di suddivisione non possono essere ricavati nello stesso flusso che li utilizza come suddivisioni; è necessario creare due flussi: uno che ricava il campo di suddivisione e uno che utilizza il campo come suddivisioni.

Riempimento

- Supporta lo stesso insieme di funzioni supportato dal nodo Ricava.

Discretizzazione

La seguente funzionalità non è supportata.

- Discretizzazione ottimale
- Ranghi
- Tiles -> Tagliamento: Sum dei valori
- Tiles -> Ties: Keep in current e Assign casualmente
- Tiles -> Custom N: Valori su 100, e qualsiasi valore N dove 100% N non è uguale a zero.

Analisi RFM

- L'opzione Mantieni in corrente per la gestione dei pari merito non è supportata. I punteggi monetari, l'attualità RFM e la frequenza non sempre corrispondono a quelli calcolati da Modeler dagli stessi dati. Gli intervalli di punteggio saranno gli stessi ma le assegnazioni del punteggio (numeri bin) potrebbero differire di uno.

Grafici

Tutti i nodi grafico sono supportati.

Modellazione

Sono supportati i seguenti nodi di modellazione: Serie temporali, TCM, Tree-AS, C&R Tree, Quest, CHAID, Lineare, Linear-AS, Rete neurale, GLE, LSVM, TwoStep-AS, Random Trees, STP, e Regole di associazione. Di seguito sono riportate ulteriori note sulla funzionalità di tali nodi.

Lineare

Quando si creano modelli su dati di grosse dimensioni, in genere si modifica l'obiettivo su Dataset di grandi dimensioni o si specificano le suddivisioni.

- L'addestramento continuato dei modelli PSM esistenti non è supportato.
- L'obiettivo di creazione modello Standard è consigliato solo se sono definiti i campi di suddivisione in modo che il numero di record in ciascuna suddivisione non sia troppo elevato, dove la definizione di "troppo elevato" dipende dalla potenza dei singoli nodi nel cluster Hadoop. Di contro, è necessario anche assicurarsi che le suddivisioni non siano definite così finemente da essere presenti troppi pochi record per creare un modello.
- L'obiettivo Aumento non è supportato.
- L'obiettivo Bagging non è supportato.
- L'obiettivo Dataset di grandi dimensioni non è consigliato quando sono presenti pochi record; spesso non verrà creato alcun modello o ne verrà creato uno ridotto.
- La preparazione automatica dei dati non è supportata. Ciò può comportare problemi nel tentare di creare un modello basato su dati privi di numerosi valori, di norma, immessi nell'ambito della preparazione automatica dei dati. Una soluzione alternativa è utilizzare un modello della struttura oppure una rete neurale impostata su Avanzata per assegnare i valori mancanti selezionati.
- La statistica di accuratezza non viene calcolata per i modelli di suddivisione.

Rete neurale

Quando si creano modelli su dati di grosse dimensioni, in genere si modifica l'obiettivo su Dataset di grandi dimensioni o si specificano le suddivisioni.

- L'addestramento continuato dei modelli standard o PSM esistenti non è supportato.
- L'obiettivo di creazione modello Standard è consigliato solo se sono definiti i campi di suddivisione in modo che il numero di record in ciascuna suddivisione non sia troppo elevato, dove la definizione di "troppo elevato" dipende dalla potenza dei singoli nodi nel cluster Hadoop. Di contro, è necessario anche assicurarsi che le suddivisioni non siano definite così finemente da essere presenti troppi pochi record per creare un modello.
- L'obiettivo Aumento non è supportato.
- L'obiettivo Bagging non è supportato.
- L'obiettivo Dataset di grandi dimensioni non è consigliato quando sono presenti pochi record; spesso non verrà creato alcun modello o ne verrà creato uno ridotto.
- Quando mancano numerosi valori nei dati, utilizzare l'impostazione Avanzata per assegnare i valori mancanti.
- La statistica di accuratezza non viene calcolata per i modelli di suddivisione.

C&R Tree, CHAID e Quest

Quando si creano modelli su dati di grosse dimensioni, in genere si modifica l'obiettivo su Dataset di grandi dimensioni o si specificano le suddivisioni.

- L'addestramento continuato dei modelli PSM esistenti non è supportato.
- L'obiettivo di creazione modello Standard è consigliato solo se sono definiti i campi di suddivisione in modo che il numero di record in ciascuna suddivisione non sia troppo elevato, dove la definizione di "troppo elevato" dipende dalla potenza dei singoli nodi nel cluster Hadoop. Di contro, è necessario anche assicurarsi che le suddivisioni non siano definite così finemente da essere presenti troppi pochi record per creare un modello.
- L'obiettivo Aumento non è supportato.
- L'obiettivo Bagging non è supportato.
- L'obiettivo Dataset di grandi dimensioni non è consigliato quando sono presenti pochi record; spesso non verrà creato alcun modello o ne verrà creato uno ridotto.
- Le sessioni interattive non sono supportate.
- La statistica di accuratezza non viene calcolata per i modelli di suddivisione.
- Quando è presente un campo suddivisione, i modelli della struttura ad albero creati in locale in Modeler sono leggermente differenti dai modelli della struttura ad albero creati da Server analitico e pertanto producono punteggi differenti. Gli algoritmi in entrambi i casi sono validi: gli algoritmi utilizzati da Server analitico sono semplicemente più recenti. Partendo dal presupposto che gli algoritmi della struttura ad albero tendono ad avere troppe regole euristiche, la differenza tra i due componenti è normale.

Calcolo del punteggio del modello

Tutti i modelli supportati per la modellazione sono supportati anche per il calcolo del punteggio.

Inoltre, i nugget del modello creati in locale per i seguenti nodi sono supportati per il calcolo del punteggio: C&RT, Quest, CHAID, Lineare e Rete neurale (indipendentemente che il modello sia standard, con boosting, con bagging o per dataset di grandi dimensioni), Regressione, C5.0, Logistica, Genlin, GLMM, Cox, SVM, Rete di Bayes, TwoStep, KNN, Elenco di decisioni, Discriminante, Autoapprendimento, Rilevamento anomalie, Apriori, Carma, Medie K, Kohonen, R e Text Mining.

- Non verranno calcolate propensioni grezze o corrette. Come soluzione temporanea è possibile ottenere lo stesso effetto calcolando manualmente la propensione grezza utilizzando un nodo Ricava con la seguente espressione: `if 'valore-previsto' == 'valore-di-interesse' then 'prob-di-quel-valore' else 1-'prob-di-quel-valore' endif`

R

La sintassi R nel nugget deve essere costituita da operazioni record-at-a-time.

Output

I nodi Matrice, Analisi, Verifica dati, Trasformazioni, Calcola globali, Statistiche, Medie e Tabelle sono supportati. Di seguito sono riportate ulteriori note sulla funzionalità dei nodi supportati.

Verifica dati

Il nodo Verifica dati non può produrre la modalità per i campi continui.

Medie

Il nodo Medie non può produrre un errore standard o un intervallo di confidenza pari al 95%.

Tabella

Il nodo Tabella è supportato mediante la scrittura di un'origine dati Server analitico temporanea che contiene i risultati di operazioni upstream. Il nodo Tabella scorre attraverso il contenuto di tale origine dati.

Esporta

Un flusso può iniziare con un nodo origine Server analitico e terminare con un nodo di esportazione diverso dal nodo di esportazione di Server analitico ma i dati si sposteranno da HDFS su SPSS Modeler Server, ed infine sulla posizione di esportazione.

Pushback a HCatalog/Hive

Pushback a HCatalog/Hive

Quando si lavora con i dati in una tabella Hive partizionata, è possibile strutturare il flusso Modeler per spingere la selezione delle partizioni desiderate in Hive.

1. Iniziare il tuo flusso con un nodo di origine Server analitico che rimanda all'origine dati HCatalog/Hive .
2. Collegarsi ad altri nodi come normalmente si farebbe.

Le seguenti sezioni forniscono Server analitico operazioni e funzioni che supportano la generazione SQL.

Server analiticooperazioni

Tabella 1. Operazioni su record				
IBM SPSS Modeler operazioni che supportano la generazione SQL		IBM SPSS Analytic Serveroperazioni	BigSQL	Hive
Seleziona				Sì *
Esempio			Sì *	Sì *
Ordinamento			Sì	Sì
Bilanciare				
Distinto			Sì	Sì
Aggregato	SOMMA		Sì	Sì
	MEDIA		Sì	Sì
	MIN		Sì	Sì
	MAX		Sì	Sì
	DEVS		Sì	Sì
	CONTEGGIO		Sì	Sì
	ContoTNON		Sì	Sì
	Varianza		Sì	Sì
	@	secondmaggiore	N	

Tabella 1. Operazioni su record (Continua)				
IBM SPSS Modeler operazioni che supportano la generazione SQL		IBM SPSS Analytic Serveroperazioni	BigSQL	Hive
	@	tredmaggiore	N	
	@	sumOfQuadrato		
Aggregazione RFM				N
Unione	interno		Sì	Sì
	esterno		Sì	Sì
	anti		N	N
	@	join.cartesian	N	
Accoda			Sì	Sì
Streaming di TS			N	N
Streaming di TCM			N	N

@ - Non esiste una funzione corrispondente in SPSS Modeler.

* - Il nodo supporta molte funzioni, alcune delle quali supportano il pushback.

Tabella 2. Operazioni su campi			
IBM SPSS Modeler operazioni che supportano la generazione SQL	IBM SPSS Analytic Serveroperazioni	Push back BigSQL	Hive pushback
Immettere			Sì
Filtro		Sì	Sì
Ricava		Sì *	Sì *
Riempimento			Sì
Ricodifica		Sì	Sì
Discretizzazione			N
Analisi RFM			N
Insieme			N
Partizione			N
Imposta indicatore			Sì
Riorganizza			N
Riordina campi			Sì
Riproiezione			N
Intervalli di tempo			N

* - Il nodo supporta molte funzioni e alcuni di essi supportano il pushback.

<i>Tabella 3. Operazioni di esportazione</i>			
IBM SPSS Modeler operazioni che supportano la generazione SQL	IBM SPSS Analytic Serveroperazioni	Push back BigSQL	Hive pushback
Memorizzazione automatica			

<i>Tabella 4. Operazioni del grafico</i>			
IBM SPSS Modeler operazioni che supportano la generazione SQL	IBM SPSS Analytic Serveroperazioni	Push back BigSQL	Hive pushback
Lavagna grafica			Sì *
Grafico			N
Multiplot			N
Grafico temporale			N
Distribuzione			Sì
Istogramma			N
Raccolta			N
web			Sì
Valutazione			N
Visualizzazione mappa			N
E-Plot			N

* - Il nodo supporta molte funzioni e alcuni di essi supportano il pushback.

<i>Tabella 5. Operazioni di Nugget</i>			
IBM SPSS Modeler operazioni che supportano la generazione SQL	IBM SPSS Analytic Serveroperazioni	Push back BigSQL	Hive pushback
GLE			N
Lineare-AS			N
LSVM			N
AS-Albero			N
Serie temporali			N
TCM			N
STP			N
TwoStep-AS			N
Regole di associazione			N

<i>Tabella 6. Operazioni di output</i>			
IBM SPSS Modeler operazioni che supportano la generazione SQL	IBM SPSS Analytic Serveroperazioni	Push back BigSQL	Hive pushback
Tabella			N
Matrice			Sì

<i>Tabella 6. Operazioni di output (Continua)</i>			
IBM SPSS Modeler operazioni che supportano la generazione SQL	IBM SPSS Analytic Server operazioni	Push back BigSQL	Hive pushback
Analisi			N
Verifica dati			N
Trasforma			Sì
Statistiche			N
Medie			N
Relazione			Sì *
Imposta valori globali			Sì

* - Il nodo supporta molte funzioni e alcuni di essi supportano il pushback.

Funzioni di Server analitico

<i>Tabella 7. Funzioni aritmetiche</i>			
IBM SPSS Modeler funzioni aritmetiche	IBM SPSS Analytic Server funzioni aritmetiche	Push back BigSQL	Hive pushback
+	+	Sì	Sì
-	-	Sì	Sì
*	*	Sì	Sì
/	/	Sì	Sì
@	%	Sì	
@	^	Sì	
@	isNull	Sì	

@ - Non esiste una funzione corrispondente in SPSS Modeler.

<i>Tabella 8. Funzioni di bit</i>			
IBM SPSS Modeler bit funzioni	IBM SPSS Analytic Server bit funzioni	Push back BigSQL	Hive pushback
@	bitAnd	Sì	
@	bitAndEqualZero	Sì	
@	bitAndNon	Sì	
@	bitAndNotEqualZero	Sì	
@	bitNot	Sì	
@	bitOr	Sì	
@	bitXor	Sì	
integer_bitcount	Conteggio intBit	N	N
integer_leastbit	Bit intLeast	N	N

Tabella 8. Funzioni di bit (Continua)

IBM SPSS Modeler bit funzioni	IBM SPSS Analytic Server bit funzioni	Push back BigSQL	Hive pushback
integer_lunghezza	intLength	N	N
testbit	testBit	N	N

@ - Non esiste una funzione corrispondente in SPSS Modeler.

Tabella 9. Funzioni carattere

IBM SPSS Modeler Funzioni carattere	IBM SPSS Analytic Server Funzioni carattere	Push back BigSQL	Hive pushback
Codice isAlpha	Codice isAlpha	Sì	
Codice isNumber	Codice isNumber	Sì	
Codice isLower	Codice isLower	Sì	
Codice isUpper	Codice isUpper	Sì	
unicode_char	unicodeChar	N	N
unicode_value	unicodeValue	N	

Tabella 10. funzioni di confronto

IBM SPSS Modeler funzioni di confronto	IBM SPSS Analytic Server funzioni di confronto	Push back BigSQL	Hive pushback
=	==	Sì	Sì
>=	>=	Sì	Sì
>	>	Sì	Sì
<=	<=	Sì	Sì
<	<	Sì	Sì
/=	!=	Sì	Sì
non	!	Sì	Sì
e	e	Sì	Sì
oppure	oppure	Sì	Sì
count_uguale	countEqual	Sì	Sì
count_greater_than	countGreaterThan	Sì	Sì
consiglia_less_than	countLessThan	Sì	Sì
count_no_uguale	countNotuguale	Sì	Sì
contot_nulli	countnulls	Sì	Sì
first_index	firstindex	N	N
first_non_null_index	firstnullindex	Sì	Sì
last_index	LASTINDEX	N	N

Tabella 10. funzioni di confronto (Continua)

IBM SPSS Modeler funzioni di confronto	IBM SPSS Analytic Server funzioni di confronto	Push back BigSQL	Hive pushback
last_non_null_index	lastnonnullindex	Sì	Sì
vale_at	valueAt	Sì	Sì
@	pasticcio di carne	N	

@ - Non esiste una funzione corrispondente in SPSS Modeler.

Tabella 11. Funzioni di conversione

IBM SPSS Modeler funzioni di conversione	IBM SPSS Analytic Server funzioni di conversione	Push back BigSQL	Hive pushback
to_Int	TOINT	Sì	Sì
to_reale	toReal	Sì	Sì
to_stringa	toString	N	Sì
to_data	toDate	Sì	Sì
to_ora	toTime	N	N
to_timestamp	toTimestamp	N	N

Tabella 12. Funzioni di data e ora

Funzioni data e ora IBM SPSS Modeler date and time functions	Funzioni data e ora IBM SPSS Analytic Server date and time functions	Push back BigSQL	Hive pushback
datetime_now	ora	Sì	Sì
Oggi	oggi	N	N
@	formato	N	N
datetime_date_name (DAY)	dayName	N	N
datetime_month_name (MESE)	monthName	N	N
datetime_date(ELEMENTO)	data (Numero)	Sì	N
datetime_date(ELEMENTO)	data (data/ora)	Sì	N
@	data (stringa, formato, locale)	N	
datetime_date(ANNO,MESE,GIORNO)	DATE(ANNO,MESE,GIORNO)	Sì	Sì
datetime_date(ELEMENTO)	ora (Numero)	N	N
datetime_date(ELEMENTO)	ora (ora)	Sì	N
@	ora (stringa, formato, locale)	N	N

Tabella 12. Funzioni di data e ora (Continua)

Funzioni data e ora IBM SPSS Modeler date and time functions	Funzioni data e ora IBM SPSS Analytic Server date and time functions	Push back BigSQL	Hive pushback
datetime_time(ORA,MINUTO,SECONDO)	ora (ora, minuto, secondo)		
	timefromtimestamp		Sì
datetime_timestamp (elemento)	registrazione data/ora (numero)	N	N
datetime_timestamp(DATA, ORA)	registrazione data / ora (data, ora)	Sì	N
@	data/ora (stringa, formato, locale)	N	N
datetime_timestamp(ANNO,MESE,GIORNO,ORA,MINUTO,SECONDO)	data/ora (y, m, d, h, m)		N
datetime_year(DATA)	anno	Sì	N
datetime_month(DATA)	mese	Sì	N
datetime_day(DATA)	giorno	Sì	N
datetime_weekday(DATA)	dayOfWeek	Sì	N
datetime_hour(ORA)	ore	Sì	N
datetime_minute(ORA)	minuti	Sì	N
datetime_second(ORA)	secondi	Sì	N
@	millisecondi	Sì	
date_before	dateBefore	Sì	N
time_before	timeBefore	Sì	N
date_days_difference	daysDifference	Sì	N
date_weeks_difference	weeksDifference	Sì	N
date_months_difference	monthsDifference	Sì	N
date_years_difference	yearsDifference	Sì	N
time_hours_difference	hoursDifference	Sì	N
time_mins_difference	minutesDifference	Sì	N
time_secs_difference	secondsDifference	Sì	N
time_in_hours	timeInOre	Sì	N
time_in_secondi	timeInSecondi	Sì	N
date_in_days	dateIngiorni	Sì	N
date_in_weeks	dateInSettimane	Sì	N
date_in_months	dateInMesi	Sì	N
date_in_years	dateInanni	Sì	N

@ - Non esiste una funzione corrispondente in SPSS Modeler.

<i>Tabella 13. Funzioni di elenco</i>			
IBM SPSS Modeler funzioni di elenco	IBM SPSS Analytic Server funzioni di elenco	Push back BigSQL	Hive pushback
@	argMax		
@	argMin		
@	concatList		
@	gruppo		
@	indexMax		
@	indexMin		
@	associazione		
@	partizione		
@	reduceLeft		
@	inversione		
@	sortList		

@ - Non esiste una funzione corrispondente in SPSS Modeler.

<i>Tabella 14. Funzioni di informazione</i>			
IBM SPSS Modeler funzioni di informazione	IBM SPSS Analytic Server funzioni di informazione	Push back BigSQL	Hive pushback
is_data	isDate	Sì	Sì
is_datetime	isDateOra	Sì	Sì
is_intero	isInt	Sì	Sì
is_number	isNumber		N
is_reale	isReal	Sì	Sì
is_string	isString		N
is_ora	isTime	Sì	Sì
is_timestamp	isTimestamp	Sì	Sì

<i>Tabella 15. Funzioni matematiche</i>			
IBM SPSS Modeler funzioni matematiche	IBM SPSS Analytic Server funzioni matematiche	Push back BigSQL	Hive pushback
abs	abs	Sì	Sì
div	divWhole	N	N
@	e	Sì	
exp	exp	Sì	Sì

Tabella 15. Funzioni matematiche (Continua)

IBM SPSS Modeler funzioni matematiche	IBM SPSS Analytic Server funzioni matematiche	Push back BigSQL	Hive pushback
frattof	frac	Sì	Sì
@	ln	Sì	
log10	log10	Sì	Sì
pi	pi	Sì	Sì
casuale	casuale	Sì	Sì
@	realToInt	Sì	
arrotondare	arrotondare	Sì	Sì
firmare	firmare	Sì	Sì
quadrato	quadrato	Sì	Sì

@ - Non esiste una funzione corrispondente in SPSS Modeler.

Tabella 16. funzioni di probabilità

IBM SPSS Modeler funzioni di probabilità	IBM SPSS Analytic Server funzioni di probabilità	Push back BigSQL	Hive pushback
cdf_chisquare	cdfChiQuadrato	N	N
cdf_f	cdfF	N	N
cdf_normale	cdfNormal	N	N
cdf_t	cdfT	N	N

Tabella 17. Funzioni statistiche

IBM SPSS Modeler Funzioni statistiche	IBM SPSS Analytic Server Funzioni statistiche	Push back BigSQL	Hive pushback
max_n	maxN	Sì	Sì
significa__n	meanN	Sì	Sì
min_n	minN	Sì	Sì
sdev_n	sDevN	N	N
sum_n	sumN	N	N

Tabella 18. Funzioni trigonometriche

IBM SPSS Modeler funzioni trigonometriche	IBM SPSS Analytic Server funzioni trigonometriche	Push back BigSQL	Hive pushback
SIN	SIN	Sì	Sì
SINH	SINH	Sì	Sì
lattuga romana	lattuga romana	Sì	Sì

Tabella 18. Funzioni trigonometriche (Continua)

IBM SPSS Modeler funzioni trigonometriche	IBM SPSS Analytic Server funzioni trigonometriche	Push back BigSQL	Hive pushback
COSH	COSH	Sì	Sì
TAN	TAN	Sì	Sì
TANH	TANH	Sì	Sì
arcina	arcina	Sì	Sì
arcsinh	arcsinh	Sì	Sì
arccos	arccos	Sì	Sì
arccosh	arccosh	Sì	Sì
arctan	arctan	Sì	Sì
arctan2	arctan2	Sì	N
arctanh	arctanh	Sì	Sì
@	toDegrees	Sì	
@	toRadians	Sì	

@ - Non esiste una funzione corrispondente in SPSS Modeler.

Tabella 19. Funzioni di stringa

IBM SPSS Modeler funzioni stringa	IBM SPSS Analytic Server funzioni stringa	Push back BigSQL	Hive pushback
allbutprima	allButPrimo	Sì	Sì
allbutlast	allButUltimo	Sì	Sì
alphabefore	alphaBefore	Sì	Sì
@	charAt	Sì	
@	comparare	N	
@	conteggio	N	
@	detectLanguage	N	
endstring	endString	Sì	Sì
isstartstring	Stringa isStart	Sì	Sì
isendstring	Stringa isEnd	Sì	Sì
ismidstring	Stringa isMid	N	Sì
isSubstring	isSubstring	Sì	Sì
issubstring_count	isSubStringCount	N	N
issubstring_lim	isSubStringLim	N	N
@	jsonPath	N	
@	concatenazione	Sì	

Tabella 19. Funzioni di stringa (Continua)

IBM SPSS Modeler funzioni stringa	IBM SPSS Analytic Server funzioni stringa	Push back BigSQL	Hive pushback
Ultimo	Ultimo	Sì	Sì
lunghezza	lunghezza	Sì	Sì
@	Byte lengthIn	Sì	Sì
locchar	locChar	Sì	
locchar_back	locCharIndietro	Sì	Sì
uppertolower	inferiore	Sì	
trimstart	lTrim	Sì	Sì
sostituire	sostituire	Sì	N
replicare	replicare	Sì	Sì
trimend	rTrim	Sì	Sì
skipchar	skipChar	N	N
skipchar_back	skipCharIndietro	N	N
SoundEx	soundEx	Sì	Sì
soundex_difference	Differenza soundEx	Sì	Sì
startstring	startString	Sì	Sì
stripchar	stripChar	Sì	
strmember	strMember	Sì	
sottostringa	subString	Sì	N
substring_between	subStringTra	Sì	N
ritagliare	ritagliare	Sì	Sì
lowertoupper	superiore	Sì	

@ - Non esiste una funzione corrispondente in SPSS Modeler.

Tabella 20. Funzioni geospaziali

IBM SPSS Modeler funzioni geospaziali	IBM SPSS Analytic Server funzioni geospaziali	Push back BigSQL	Hive pushback
close_to	closeTo	N	N
crosses	crosses	N	N
@	intersecare	N	
@	toccare	N	
sovrapporre	sovrapporre	N	N
all'interno di	all'interno di	N	N
@	contenere	N	

Tabella 20. Funzioni geospaziali (Continua)

IBM SPSS Modeler funzioni geospaziali	IBM SPSS Analytic Server funzioni geospaziali	Push back BigSQL	Hive pushback
@	northOf	N	
@	southOf	N	
@	eastOf	N	
@	westOf	N	
@	baricentro	N	
area	area	N	N
num_points	numPoints	N	N

@ - Non esiste una funzione corrispondente in SPSS Modeler.

Capitolo 3. Risoluzione dei problemi

In questa sezione vengono descritti alcuni problemi di utilizzo comune e come è possibile risolverli.

origini dati

I filtri definiti sulle colonne partizionate nelle origini dati HCatalog non sono onorati

Questo è un problema visto in alcune versioni di Hive, e può essere visto nelle seguenti situazioni.

- Se si definisce un sorgente dati HCatalog e si specifica un filtro nella definizione dell'origine dati.
- Se si crea un flusso Modeler con un nodo Filter che fa riferimento alla colonna della tabella partizionata.

Il workaround è quello di aggiungere un nodo Deriva al flusso Modeler che crea un nuovo campo con valori pari alla colonna partizionata. Il nodo Filter dovrebbe fare riferimento a questo nuovo campo.

Oracle NoSQL

Si sono verificati errori "Esecuzione non riuscita" durante la connessione a un'origine dati Oracle NoSQL

Il problema è dovuto al gestore di archiviazione HiveKVStorageHandler.jar obsoleto.

Deve essere utilizzato un gestore di archiviazione aggiornato. Il file aggiornato può essere trovato a 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. Copiare il file JAR nella directory Hive {HIVE_HOME}/auxlib e la directory Server analitico {AS_ROOT}/ae_wlpserver/usr/servers/aeserver/apps/AE_BOOT.war/WEB-INF/lib.
2. Eseguire {AS_ROOT}/bin/hdfsUpdate.sh per propagare le modifiche a HDFS.
3. Riavviare Server analitico per rendere effettive le modifiche.

Nota: La classe del gestore di memorizzazione `oracle.kv.hadoop.hive.table.TableStorageHandler` è consigliata quando viene utilizzato il database Oracle NoSQL 3.0. La classe richiede agli utenti di organizzare i dati con una metafora della tabella.

Informazioni particolari

Queste informazioni sono state sviluppate per prodotti e servizi offerti negli Stati Uniti. Questo materiale potrebbe essere disponibile da IBM in altre lingue. Tuttavia, per potervi accedere, potrebbe essere richiesto di possedere una copia del prodotto o una versione del prodotto in tale lingua.

È possibile che IBM non offra in altri paesi i prodotti, i servizi o le funzioni illustrati in questa documentazione. Consultare il proprio rappresentante locale IBM per informazioni sui prodotti ed i servizi attualmente disponibili nella propria zona. Ogni riferimento a prodotti, programmi o servizi IBM non implica che possano essere utilizzati solo quei prodotti, programmi o servizi IBM. In sostituzione a quelli forniti da IBM, è possibile usare prodotti, programmi o servizi funzionalmente equivalenti che non comportino violazione dei diritti di proprietà intellettuale o di altri diritti di IBM. Tuttavia, è responsabilità dell'utente valutare e verificare il funzionamento di qualsiasi prodotto, programma o servizio non IBM.

IBM potrebbe avere brevetti o domande di brevetti in corso relativi agli argomenti discussi nella presente pubblicazione. La fornitura del presente documento non concede alcuna licenza a tali brevetti. È possibile inviare per iscritto richieste di licenze a:

Director of Commercial Relations
IBM Italia
Schoenaicher Str. 220
D-7030 Boeblingen
US

Per richieste di licenze relative ad informazioni double-byte (DBCS) contattare il Dipartimento di Proprietà Intellettuale IBM nel proprio paese o inviare richieste per iscritto a:

Licenza per la proprietà intellettuale
Legge sulla proprietà intellettuale e legale
IBM Japan, Ltd.
19-21, Nihonbashi-Hakozakicho, Chuo-ku
Tokyo 103-8510, Giappone

INTERNATIONAL BUSINESS MACHINES CORPORATION FORNISCE QUESTA PUBBLICAZIONE NELLO STATO IN CUI SI TROVA, SENZA ALCUNA GARANZIA, ESPLICITA O IMPLICITA, IVI INCLUSE EVENTUALI GARANZIE DI COMMERCIALIZZABILITÀ E DI IDONEITÀ AD UNO SCOPO PARTICOLARE. Alcune giurisdizioni non consentono la rinuncia a garanzie esplicite o implicite in determinate transazioni; quindi la presente dichiarazione potrebbe non essere a voi applicabile.

Queste informazioni potrebbero includere imprecisioni tecniche o errori tipografici. Le modifiche alle presenti informazioni vengono effettuate periodicamente; tali modifiche saranno incorporate nelle nuove edizioni della pubblicazione. IBM si riserva il diritto di apportare miglioramenti e/o modifiche al prodotto o al programma descritto nella presente pubblicazione in qualsiasi momento e senza alcun preavviso.

Qualsiasi riferimento contenuto nelle presenti informazioni a siti web non-IBM è fornito solo per praticità e non costituisce in alcun modo un'approvazione ufficiale di tali siti web. I materiali contenuti in tali siti Web non fanno parte di questo prodotto IBM e l'utente si assume ogni rischio relativo al loro utilizzo.

IBM può utilizzare o distribuire qualsiasi informazione fornita dall'utente nel modo che ritenga più idoneo senza incorrere in alcun obbligo nei confronti dell'utente stesso.

Coloro che detengano la licenza su questo programma e desiderano avere informazioni su di esso allo scopo di consentire: (i) uno scambio di informazioni tra programmi indipendenti ed altri (compreso questo) e (ii) l'utilizzo reciproco di tali informazioni, dovrebbe rivolgersi a:

Director of Commercial Relations
IBM Italia
Schoenaicher Str. 220

Tali informazioni potrebbero essere disponibili secondo termini e condizioni appropriati compreso, in alcuni casi, il pagamento di un corrispettivo.

Il programma concesso in licenza e descritto nella presente documentazione e tutto il materiale su licenza ad esso relativo vengono forniti da IBM nel rispetto dei termini dell'IBM Customer Agreement, dell'IBM International Program License Agreement o di qualunque altro accordo equivalente tra le parti.

I dati prestazione e gli esempi client citati vengono presentati solo per fini illustrativi. I risultati della prestazione reali possono variare a seconda delle configurazioni specifiche e delle condizioni d'operazione.

Le informazioni relative ai prodotti non-IBM sono state ottenute dai fornitori di tali prodotti, dai relativi annunci pubblicati o dalle relative fonti pubbliche. IBM non ha testato quei prodotti e non può garantire l'accuratezza delle prestazioni, la compatibilità o qualsiasi altra dichiarazione relativa a prodotti non IBM. Eventuali commenti relativi alle prestazioni dei prodotti non-IBM devono essere indirizzati ai fornitori di tali prodotti.

Le dichiarazioni riguardanti in generale l'orientamento o gli scopi futuri di IBM è soggetta a modifica o a ritiro senza preavviso.

Tutti i prezzi IBM devono considerarsi prezzi al dettaglio suggeriti da IBM, sono attuali e possono essere soggetti a modifica senza preavviso. I prezzi al fornitore possono variare.

Queste informazioni vengono fornite solo per scopi di pianificazione. Le presenti informazioni sono soggette a modifiche prima che i prodotti descritti siano resi disponibili.

Questa pubblicazione contiene esempi di dati e prospetti utilizzati quotidianamente nelle operazioni aziendali. Per meglio illustrarli, tali esempi possono contenere nomi di persone, società, marchi e prodotti. Tutti questi nomi sono fittizi e qualsiasi somiglianza con persone reali o aziende business è puramente casuale. *to actual people or business enterprises is entirely coincidental.*

LICENZA DI COPYRIGHT:

Questa pubblicazione contiene esempi di dati e prospetti utilizzati quotidianamente nelle operazioni aziendali. Per meglio illustrarli, tali esempi possono contenere nomi di persone, società, marchi e prodotti. Tutti questi nomi sono fittizi e qualsiasi somiglianza con persone reali o aziende business è puramente casuale. *to actual people or business enterprises is entirely coincidental.*

Ogni copia o parte di questi programmi di esempio o dei lavori derivati deve includere le informazioni sul copyright, come di seguito riportato:

© IBM 2020. Le porzioni di questo codice derivano da IBM Corp. Programmi di esempio.

© Copyright IBM Corp. 1989 - 2020. Tutti i diritti riservati.

Marchi

IBM, il logo IBM e ibm.com sono marchi o marchi registrati di International Business Machines Corp., registrati in molte giurisdizioni in tutto il mondo. Altri nomi di prodotti e servizi possono essere marchi commerciali di IBM o di altre aziende. Un elenco aggiornato di marchi IBM è disponibile sul web in "Copyright and trademark information" su www.ibm.com/legal/copytrade.shtml.

Adobe, il logo Adobe, PostScript e il logo PostScript sono marchi o marchi registrati di Adobe Systems Incorporated negli Stati Uniti e/o in altri paesi.

IT Infrastructure Library è un marchio registrato di Central Computer and Telecommunications Agency, ora parte di Office of Government Commerce.

Intel, Intel logo, Intel Inside, Intel Inside logo, Intel Centrino, Intel Centrino logo, Celeron, Intel Xeon, Intel SpeedStep, Itanium e Pentium sono marchi o marchi registrati di Intel Corporation o relative controllate negli Stati Uniti e altri paesi.

Linux è un marchio registrato di Linus Torvalds negli Stati Uniti e/o in altri paesi.

Microsoft, Windows, Windows NT e il logo Windows sono marchi di Microsoft Corporation negli Stati Uniti e/o negli altri paesi.

ITIL è un marchio registrato e un marchio comunitario registrato di The minister for the Cabinet Office ed è registrato negli U.S. Patent and Trademark Office.

UNIX è un marchio registrato di The Open Group negli Stati Uniti e in altri paesi.

Cell Broadband Engine è un marchio di Sony Computer Entertainment, Inc. negli Stati Uniti e/o in altri paesi e utilizzato su licenza.

Linear Tape-Open, LTO, il logo LTO, Ultrium ed il logo Ultrium sono marchi di HP, IBM Corp. e Quantum negli Stati Uniti e/o in altri paesi.

