InfoSphere Information Server

Importing text file metadata for Information Analyzer profiling

© 2011 IBM Corporation

This presentation explains how to import metadata from a text file so it can later be profiled in Information Analyzer.

Objectives

- Configure data source name
- Configure data store and data connection
- Define table structure in QETXT.INI
- Import metadata

Importing text file metadata for Information Analyzer profiling

The objectives of this presentation are to show the steps required to be able to profile a text file in Information Analyzer, referred to as IA. The presentation provides details on how to configure the Data Source Name, referred to as DSN, and using the IBM Text File ODBC Driver. Details on how to configure a Data Store and Data Connection within IA, and how to define the table structure in the QETXT.INI file are also included. Finally, this presentation describes how to import the metadata and how to profile the text file.

ODBC DSN for text database

- Create ODBC connection to text database on engine tier
  - Windows®:
    - Configure with ODBC Manager
  - UNIX® and Linux®:
    - Configure in $DSHOME/.odbc.ini file

```
[ODBC Data Sources]
iadb=Information Analyzer DB
cancun=oracle cancun DB
inventory=text database

[inventory]
Driver=/u1/IS85/IBM/InformationServer/Server/branded_odbc/lib/VMtxt24.so
Description=DataDirect TextFile(*.*) Driver
AllowUpdateAndDelete=0
ApplicationUsingThreads=1
CacheSize=4
CenturyBoundary=20
Database=/home/dsadm/inventory
DataFileExtension=TXT
DecimalSymbol=.
Delimiter=
FileOpenCache=0
FirstLineNames=0
IntlSort=0
ScanRows=25
TableType=Comma
UndefinedTable=GUESS
```

Directory where the QETXT.INI and data files reside

3   Importing text file metadata for Information Analyzer profiling   © 2011 IBM Corporation

IA requires the Engine layer to have a valid ODBC Data Source Name connection with the text database. If the engine is installed in Windows, you can use the 32-bit ODBC Driver Manager to create the DSN. In UNIX and Linux platforms, you configure the DSN by editing the file $DSHOME/.odbc.ini. First, add a line listing the DSN in the ODBC Data Sources section at the beginning of the file.

Next, add your entry for your text DSN. This slide shows an example text DSN. Make sure that the "Database" attribute in the DSN entry, points to the directory where the QETXT.INI and data file reside. The QETXT.INI file can be created either manually or using a wizard available in IA. Details of the QETXT.INI and data file are provided later in this presentation.

Testing ODBC DSN for text database

- UNIX and Linux - Test the DSN using the example program

```
$ cd $DSHOME
$ . ./dsenv
$ cd ../branded_odbc/example
$ ./example
DataDirect Technologies, Inc. ODBC Example Application.
Enter the data source name : inventory
Enter the user name      :
Enter the password       :
Enter SQL statements (Press ENTER to QUIT)
SQL>
```

DSN Created for Text Database

Before using the DSN within IA, test your text database DSN connection to be sure it connects successfully. In Windows, you can test the DSN by using the "Test Connection" button. In UNIX and Linux platforms, you can test the DSN by running the example program included under the branded_odbc/example folder. Branded_odbc is one level up from $DSHOME. Before you run this program, source the dsenv file. After invoking the example program you will have to provide the data source name, in this example it is "inventory". Press Enter for the user and the password. If the connection is successful, you will see an SQL prompt after entering the password. Press Enter at the prompt to exit the program. The example program must connect successfully before you can proceed to create the data source and analyze the data within IA.

Configure data source within Information Analyzer

- Add data source
- Click Home Pillar menu => Configuration => Sources
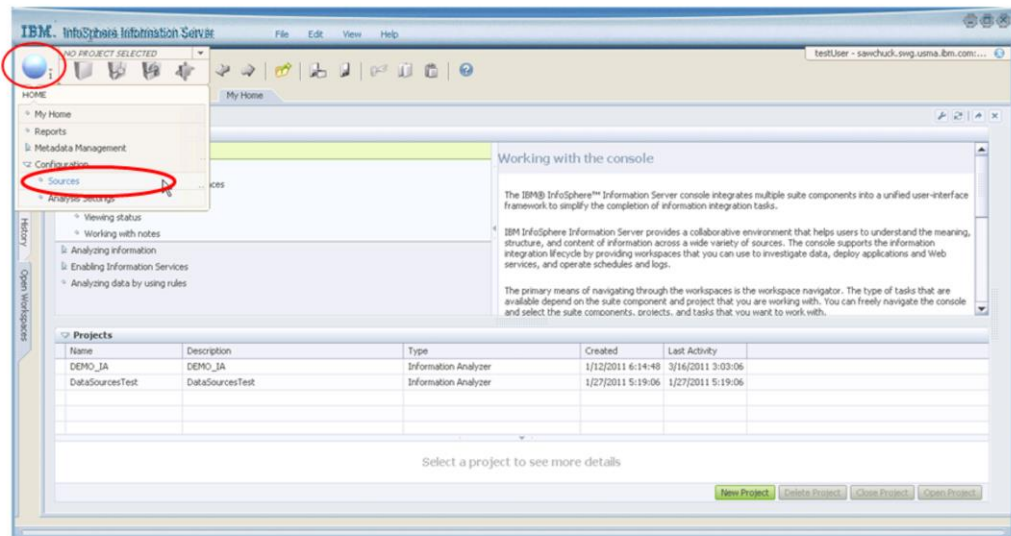
5   Importing text file metadata for Information Analyzer profiling   © 2011 IBM Corporation

To start analyzing your data in IA, create a data source to connect to your text database. To do this, open the IBM InfoSphere® Information Server Console, login with a user that has the Information Analyzer Data Administrator role and the DataStage® and QualityStage™ Administrator role. Click the Home pillar menu and click "Sources" under "Configuration". This will allow you to define a connection to the database you want to analyze.

Create new data store

- Select a host and click "New Data Store"
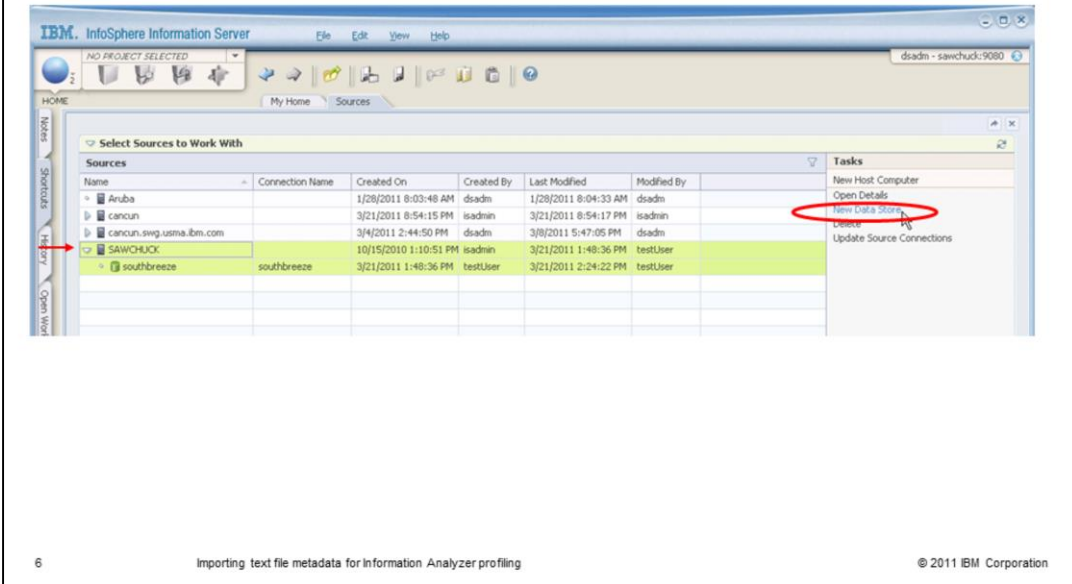
When you enter the Sources screen, you will see a list of Host Computers hosting data sources. Look at the list of Host Computers under Sources. If you see the machine you want to connect to, select it and click "New Data Store". If this is the first time connecting to a machine and it is not in the list, click New Host Computer and provide the name of the new host. In this presentation we are using the host "SAWCHUCK."

Configure data store

- Enter name for data store and data connection
- Select connector and provide database credentials

These fields are populated if the connection is successful

Define a Data Store:
Host Computer:
SAWCHUCK
Name: *
Inventory
Creation Tool:
Short Description:
Item Inventory
Long Description:

Define a Data Connection:
Name: *
inventory
Short Description:
Long Description:

Connector: *
ODBCConnector
Connection String: *
inventory
User Name:
Password:

Connect

Data Store Information:
Store Type: *
Database
Database / Path:
/home/dsadm/inventory
DBMS: *
Text
Version:
01.00.0000
Creation Tool:

Once you click "New Data Store", you are taken to the "Configure Data Store" screen. Here, you will provide the details IA needs to connect to the text file you want to analyze. Enter a name for the Data Store and the Data Connection. These names are references and do not need to match any existing resources. Then select the ODBC Connector from the Connector drop down list and select the DSN you want to connect to. Next, provide the database information. For the text database source, leave the User Name and Password fields blank. After you have entered the information, click the "Connect" button. This will validate the connectivity. If the connection is successful you will see the Data Store Information retrieved from the database. Click Save and Close. You are ready now to import metadata and start analyzing your data.

In order to import metadata from a text file, a QETXT.INI file that defines the table structure must exist in the database directory. The QETXT.INI file can be created using a text editor and it must specify the attributes of all the defined tables. Any attributes specified in the ODBC.INI or a connection string, are overridden by settings in the QETXT.INI file.

This slide displays an example of a QETXT.INI file. This file defines the structure of the items.txt file, which is a sample data file that is used throughout this presentation. The QETXT.INI must be defined as follows:

Create a Defined Tables section which lists all of the tables you are defining. Specify the text file name followed by the name you want to give the table, for example:

items.txt=ITEM

Table names can be up to 32 characters in length and cannot be the same as another defined table in the database. This name is returned by SQLTables. By default, it is the file name without its extension. For each table listed in the Defined Tables section, you must specify the text file name, the table type, whether the first line of the file contains column names, and the delimiter character.

The line FILE=items.txt specifies the text filename is items.txt. The line FLN=1 specifies the first line contains the column names, if it does not then a 0 should be specified. To define the table type, specify how the fields are separated (comma, tab, fixed, or character). For example: TT=COMMA.

Then define the fields in the table, beginning with FIELD1. For each field, specify the field name, field type, precision, scale, length, offset (for fixed tables), and date/time mask. For example, FIELD1 is named Item_ID, it is of type NUMERIC, it has a precision of 2 and a length of 8.

Sample text file

- Sample contents of items.txt file defined in sample QETXT.INI:

  "Item_ID","Type","Manufacturer"

  01,"Printer","Print Co."

  02,"Computer","ComputersRUs"

  03,"Phone","SmartPhone Inc."

- First line contains column names
- First column of each row is NUMERIC length 2
- Second and third columns are VARCHAR(20)

This slide displays a sample of an ITEM text file as defined by the QETXT.INI file. The first line contains the column names: Item_ID, Type and Manufacturer. This was designated by the use of the FLN=1 in the QETXT.INI file. The remaining lines contain the data where the first column is a number and the second and third columns are varchar. The data file should reside in the directory pointed to by the "Database" attribute in the DSN.

Using IA wizard to create QETXT.INI

- QETXT.INI created within IA using a wizard
  - http://publib.boulder.ibm.com/infocenter/iisinfsv/v8r5/topic/com.ibm.swg.im.iis.ia.administer.doc/topics/dq_importing_flat_file.html

10   Importing text file metadata for Information Analyzer profiling   © 2011 IBM Corporation
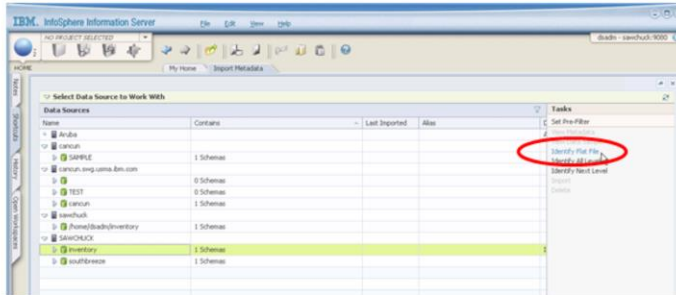
If you do not have a QETXT.INI file or do not want to create it manually, IA provides a wizard to create the QETXT.INI file. To use the wizard, highlight the data source that you want to upload the flat file to. The data source must contain at least one schema. Click Identify Flat File from the task list on the right side of the workspace. After you click Identify Flat File, a wizard is displayed.

Identify Flat File wizard
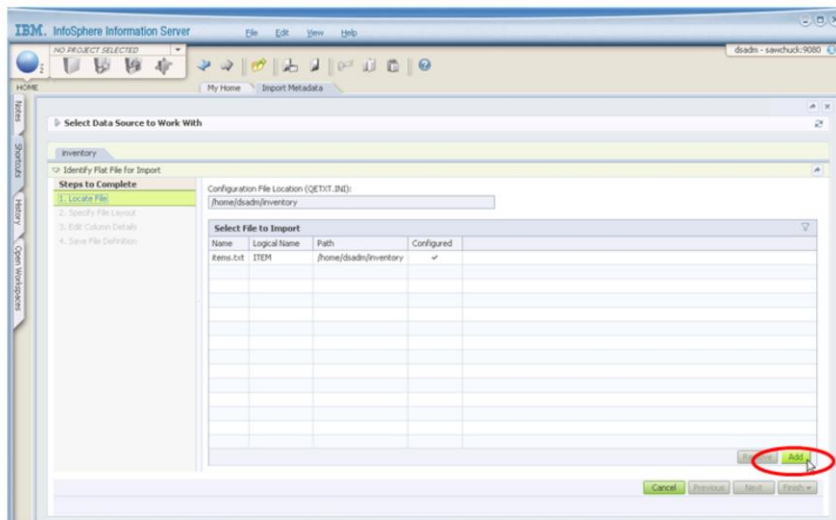
11   Importing text file metadata for Information Analyzer profiling   © 2011 IBM Corporation

In the Flat File Wizard, follow the steps to complete the task list on the left side of the screen. First, locate the file you want to import in the Select Flat File to identify workspace or click Add if you want to add a new flat file from a directory on your system. In this example, there is already one file named items.txt defined in the QETXT.INI and another file can be added. The wizard will update the existing QETXT.INI file with the new table. If a QETXT.INI file does not exist, it is created.

Testing ODBC DSN for text database

- UNIX/Linux - Test DSN using example program
$ cd $DSHOME
$ . ./dsenv
$ cd ../branded_odbc/example
$ ./example
DataDirect Technologies, Inc. ODBC Example Application.
Enter data source name : **inventory** ← DSN Created for Text Database
Enter user name   :
Enter password   :
Enter SQL statements (Press ENTER to QUIT)
SQL> **select * from ITEM** ← Table defined in QETXT.INI File
Item_ID Type Manufacturer
1   Printer Print Co.
2   ComputerComputersRUs
3   Phone SmartPhone Inc.

12   Importing text file metadata for Information Analyzer profiling   © 2011 IBM Corporation

Before attempting to import the metadata in IA, you can use the example program referenced earlier in the presentation to verify that the text file can be read using the QETXT.INI file. After connecting to the text database DSN, issue a select statement on the table defined in the QETXT.INI. In this example, it is the ITEM table. You will see the data in the items.txt file listed. If it appears to be correct, exit the example program by pressing Enter. If you are unable to connect to the DSN or view the data, there is a problem with either the DSN, the QETXT.INI file or the data itself. The example program must connect successfully before you can proceed to import the metadata and analyze the data in IA.

Import metadata

- Click Home Pillar menu => Metadata Management => Import Metadata
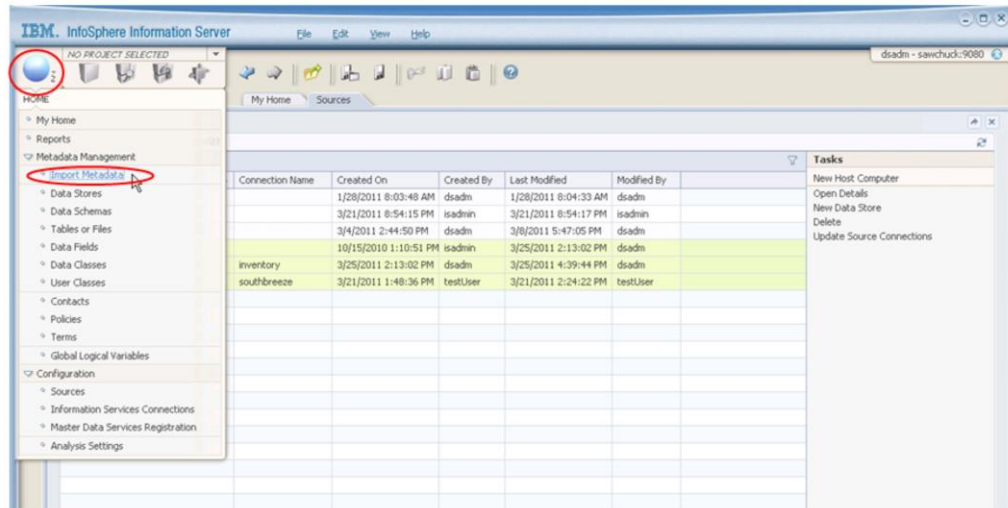
13   Importing text file metadata for Information Analyzer profiling   © 2011 IBM Corporation

Once the QETXT.INI file and data file are created and the database configuration has been verified with the example program, proceed with the metadata import. Click the Home Pillar Menu, go to "Metadata Management" and then click "Import Metadata".

Select the path under the previously defined Data Store. Click Identify Next Level. Upon completion, click OK, and expand the path node to see all discovered files and tables.

Discover tables

- Keep selection of all discovered files/tables, click "Identify Next Level" to continue to discover columns
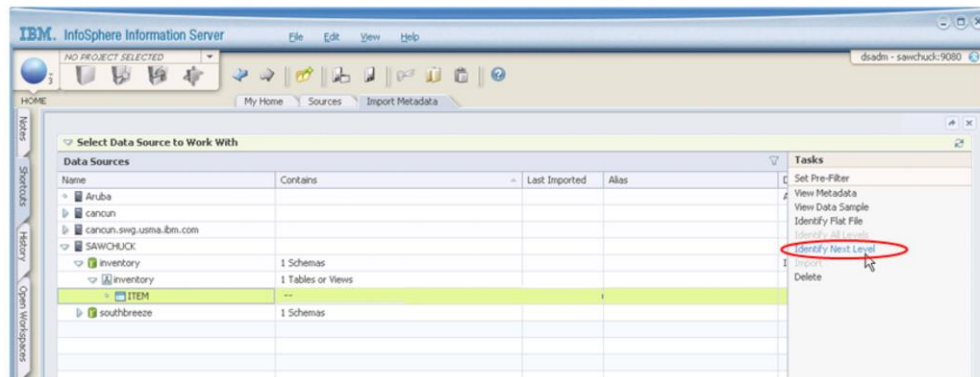
Keep the selection of all the discovered files and tables. Click "Identify Next Level" to continue to discover columns. Upon completion, click OK.

Import metadata

- Select files/tables to import metadata
- Click "Import"
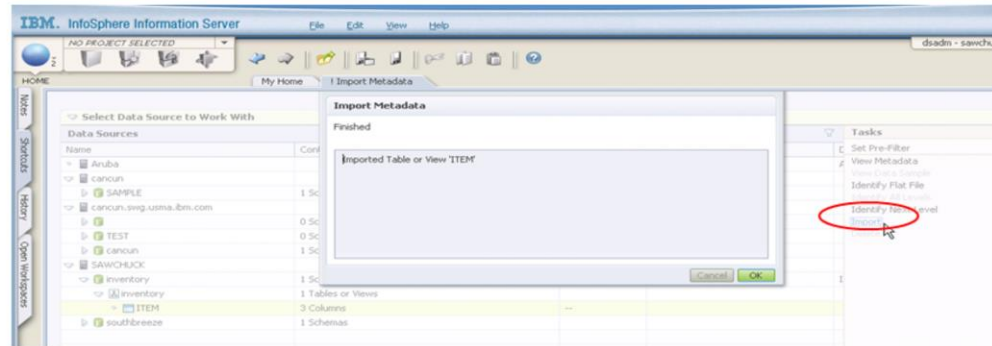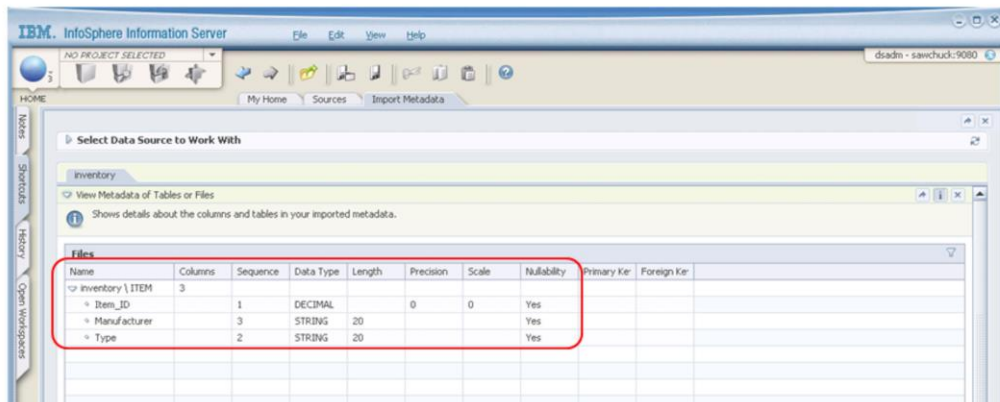
Select the files and tables whose metadata you want to import, then click "Import". You will see an Import Metadata dialog box showing the tables that were imported. Click OK to confirm.

After completing the import, the metadata can be viewed. Ensure that the Column names and types match what was defined in the QETXT.INI file. At this point, you can open a project, import the data source into the project and run column analysis on the data source.

Trademarks, disclaimer, and copyright information