

Pervasive Data Inspectors Online Help - Table of Contents

Pervasive DataTools

DataTools Data Inspectors User's Guide



Pervasive Software Inc.

12365 Riata Trace Parkway

Building B

Austin, TX 78727 USA

About This Manual

This manual leads you through the operation of the Data Inspectors user interface. The Pervasive Data Inspectors allow you to analyze data in dozens of databases, applications, or file formats. You must have a non-expired Data Inspector license to run each product.

Refer to the license.txt file in the default installation directory for disclaimers and information about trademarks and credits.

Table of Contents

- 1 [Important Installation Instructions](#)
- 1 [Introduction](#)
- 1 [Design Preferences](#)
- 1 [Source Connection Tab Overview](#)
- 1 [Source Connection](#)
- 1 [Select the Data Source Connection Type](#)
- 1 [Connecting to the Source Data](#)
- 1 [Managing the Source Connection](#)
- 1 [To Save a Source Connection](#)
- 1 [Select Data Fields and Records to Analyze](#)
- 1 [Run the Analysis](#)
- 1 [View the Results](#)
- 1 [Clear the Analysis](#)
- 1 [Save the Analysis](#)

Important Installation Instructions - Data Inspectors

The installation and operation of the Pervasive Data Inspectors requires the installation of Pervasive DataTools v9. After you order a trial or a license subscription for one or more of the Data Inspectors, follow the steps below, and in the order written, to install DataTools v9 and Data Inspector.

1. Pre-Installation Requirements

Follow ALL steps in the [System Requirements and Installation Pre-requisites](#) document. Do not install DataTools until/unless all of the specified requirements and pre-requisites are performed and/or verified.

2. Download Pervasive DataTools v9

The DataTools v9 installation EXE file can be downloaded from the [Product Downloads](#) page of the DataTools web site.

Download the installation file to your workstation's desktop.

3. Install Pervasive DataTools v9

Click on the DataTools v9 installation EXE file to start the installation wizard. Follow the prompts until the installation is finished. The installation may take several minutes.

4. Launch Pervasive DataTools v9

Launch Pervasive DataTools v9. When prompted, enter your login credentials that you used on the Pervasive DataTools web site to order your DataTools products.

This step is a good way to verify that you have installed DataTools v9 properly before installing Data Inspectors. If DataTools v9 will not start and run, do not proceed to the Data Inspector installation until the DataTools v9 installation is completed and DataTools v9 runs properly on your workstation.

5. Download Data Inspector

The DataTools v9 installation EXE file can be downloaded from the [Product Downloads](#) page of the DataTools web site.

Download the installation file to your workstation's desktop.

6. Install Data Inspector

Click on the Data Inspector installation EXE file to start the installation wizard. Follow the prompts until the installation is finished. The installation may take several minutes.

7. Launch DataTools v9

After installing Data Inspector, launch DataTools v9. The DataTools Launcher should find your Data Inspector license file via our web service and prompt you to download the file to your workstation.

8. Launch Data Inspector

Once the Data Inspector license file has been downloaded to your workstation, highlight the Data Inspector product in the DataTools Launcher's grid and then click the Launch button located under the grid. The Data Inspector should launch.

Introduction

The Data Inspectors are simple data quality analysis tools. Each Data Inspector provides an easy-to-use method of analyzing a data source.

Functionality

Data Inspectors perform the following actions:

- 1 Analyzes a data set accurately and efficiently
- 1 Generates a report and a graph of the analysis of your data
- 1 Using the report, you can isolate irregularities, quantify business rule violations, and identify null sets or values
- 1 Provides connectivity for many data sources including databases, applications, and data file formats

Features

Each Data Inspector includes the following features:

Simple three-step workflow:

1. Connect to the source data.
2. Select one or more fields/columns to analyze.
3. Click the Run button.

Predefined metrics return the results in tabular and graphic formats.

Design Preferences

The Data Inspector Design Preferences dialog box enables you to change some of the design time preferences for the following items:

- 1 AutoGen Defaults
- 1 Fonts
- 1 Logging

Autogen Defaults

The Autogen Defaults tab enables you to customize the Row Sample Size and Distinct Value Field limit.

By default, Row Sample Size is set to 10000 and Distinct Value Field limit is set to 15. This default set up is to limit the processing time and memory required to set up and run an analysis.

Fonts

The Fonts tab enables you to customize the fonts used in the Data Inspector user interface.

Logging

The Logging tab enables you to rename the log file name. By default, the log file name is Data_Investigator.log.

Source Connection Tab Overview

Data Inspectors can analyze data from many sources. Each Data Inspector supports a specific database, application, or data file format. Before you can analyze data from a particular source, you must identify the source connection type and connect to a database table, application entity, or data file.

To connect to the source data, you must specify the following:

- 1 Source connection
- 1 Data location
- 1 Data format

The next few pages of this manual provide details for each.

Source Connection

The only source connections that are available in the Data Inspectors are those whose structure is pre-defined by the existence of complete metadata. In other words, if you wish to analyze the data from a Microsoft SQL Server database table, there is a Data Inspector available. However, if you wish to analyze data from a report file or from a fixed ASCII, Btrieve or C-ISAM data file, you must first use one of the Pervasive Data Parsers to parse, unpack, and export the data to a CSV text file format.

Before you design your data analysis specification, you must connect to a data source. The following sections will guide you through the steps and the available options.

- 1 Source Connection Tab Overview
- 1 Select the Data Source Connection Type
- 1 Connecting to the Source Data
- 1 Managing the Source Connection

Select the Data Source Connection Type

1. The Source Connection tab opens upon launching the Data Inspector.

The Select Connection dialog box displays the following three tabs:

- o **Most Recently Used (Default)** - Lists the connections that have been recently used.
 - o **Factory Connections** - Lists the connections supplied with Data Inspector.
 - o **User Defined Connections** - Lists connections that you have saved.
2. Click a tab to select your connection type.

The selected connection tab provides a list of available connections in the specific Data Inspector for which you have a license.

3. Select your applicable source connection from the list.
4. Click OK to apply the selected source connection.

Alternatively, double-click the correct source connection to apply the current source connection. Data Inspector returns to the previous screen with the Source Connection tab open and the correct source connection listed in the Source Connection box.

Connecting to the Source Data

The fields that are displayed depend on the type of connection you select. However, regardless of the connection type, you should always see the following fields:

- 1 **Connector** - Lists the connectors you chose for this connection. This field cannot be edited.
- 1 **Structured Schema (Visible, but disabled)** - The Structured Schema field is visible but disabled in the Data Inspectors.
- 1 **Source File/URI** - Defines the data source location.

The label may vary based on the connection name; however, it is usually labeled Source File/URI.

Click the drop-down list to navigate to the source file.

To Connect to the Source Data

1. Click the Source File/URI arrow and navigate to the file location.

Data Inspector enables you to provide the source file / URI in any of the following ways:

- o Select a data source using the Source Connection arrow, and click the Open button (or double-click the source).
 - o You return to the previous screen with the Source Connection tab open and the correct source in the Source Connection text entry box.
 - o Enter the source location directly into the text box, but if you do, you must click the Connect button or press the Enter key for Data Inspector to connect to this source.
 - o If the source file you want to use is not accessible by directory search (that is, if it is at a remote connection accessible over a network), then specify a uniform resource identifier (URI) in the Source File/URI box either by navigating to it and selecting it in Source File/URI or by inputting it directly into the Source File/URI box.

- Optional: Enter the appropriate values into any other active text entry boxes (such as UserID and Password if they are available on the Source Connection tab.).

The steps required for these tasks varies according to the combination of source connection, data source, and profiling needs.

- Optional: If you want to customize the connector properties, change the property fields, and click Apply.

For more information on editing the connection properties, see [Managing the Source Connection](#).

- Click the Connect button if it is active.

Managing the Source Connection

In addition to the connect information on the left side of the Source Connection tab, there is a list of connector properties in the right panel. Not all files with the same connection have all of the same connection properties, and it may be easier to write a profiling specification if some of the properties are changed.

For instance, some ASCII-delimited files have a header row giving field names, while others do not. The default properties for the ASCII (Delimited) connector assume that there is no header row. If the file does have a field name header, changing the Header property so that Data Inspector will pick field names up makes defining a profiling specification easier. For the case of the Header property, double-clicking False changes the value to True. This lets Data Inspector read the field names from the header, so it uses actual field names (Name, Address, and so on) instead of the default field names (Field1, Field2, and so on) in setting up profiling metrics.

To Verify a Connection Property

- Click the Source Data Browser icon (magnifying glass) in the task bar, or select View > Source Data. This opens the Source Data Browser in which you can check the data formatting.
- Look at a sampling of the records to ensure all the fields are properly parsed.
 - If they are, go on to To Save a Source Connection.
 - If they are not, pick another connection or change the connection properties.

To Change a Connection Property

- Click in the Value field next to the Property name.

For some properties, a drop-down list arrow appears in the value cell. Other properties may require you to specify a file or enter a text value.

- Indicate the desired value. Do one of the following:
 - If there is a drop-down list, open the list and click the desired value.
 - If there is no drop-down list, enter the desired value directly in the cell.
 - If the connector property is a boolean value, you can also double-click the value to toggle it to the other state.
 - Click Apply to save your selections, or click Restore to stay with the original value.

NOTE: After making a change to one or more of the connection property options, if you do not click the Apply button, the settings revert to the previous value before your change was made.

To Save a Source Connection

- Click the Save Source Connection icon to the right of the Source Connection field. The Save Source Connection Document As window opens.
- Name the edited connection file at the File name text box and give it a location at the Save in browser box. The file is saved with an .sc.xml extension.

You can reuse the saved connection by selecting it in the User Defined Connection window.

Select Data Fields and Records to Analyze

- Connect to the data source as described in [Source Connection](#).
- Click on the Results tab.
- Click on the Source Data Sample button to open the list of data fields/columns in the data source.
- Click on the checkbox to the left of one or more data fields.
- Click on the OK button to close the dialog box.

By default, the field list will offer only the first 15 data fields from your data source.

To change the number of visible fields, follow these steps:

- Click on the View menu.
- In the View menu, click on Preferences.
- In the Preferences dialog box, click on the AutoGen Defaults tab.
- Highlight the default value (15) in the Distinct Value Field Limit box and enter a value that will display the data field(s) in the filter grid on which you wish to run the analysis.
- Click the OK button to close the Preferences dialog box.

The Data Inspector will run an analysis on the selected data field(s) and will return a table and a pie chart of Distinct Values on the field or fields.

By default, the Data Inspector will run the Distinct Values analysis on the selected fields of the first 10,000 data records.

To change the number of data records on which to run the analysis, follow these steps:

1. Click on the Source Data Sample button in the button bar.
2. In the right portion of the Data Filters dialog box, highlight the default value in the "end with:" box to a value that suits your analysis requirements.
3. Click the OK button to close the Data Filters dialog box.

Note: The Distinct Values analysis requires significant amounts of processing. The selection of a large number of data fields in a large data set may significantly increase the time to completion.

After selecting the desired data fields and data records on which to run the Distinct Values analysis, click on the Run Profile button in the button bar.

Run the Analysis

After connecting to the data source and selecting the desired data fields and data records on which to run the Distinct Values analysis, click on the Run Profile button in the button bar.

Depending on the number of data fields and data records you chose to analyze, the analysis will take a few seconds to a few minutes to complete. Note: The Distinct Values analysis requires significant amounts of processing. The selection of a large number of data fields in a large data set may significantly increase the time to completion.

When the analysis is complete, a table and a pie chart showing the results will display on the Results tab.

View the Results

After the analysis is complete, the Results tab will display two tables (on the left) and a pie chart (on the right). The following explains how to read the results.

Table Results

Upper Table

The information displayed in the upper table of the Results tab is the result of the Distinct Value analysis of the selected data source.

You will see at least two rows of information listed in the upper table. The first row contains the number of records that were analyzed in the data source. The second row contains the field name of one of the first source data field that was analyzed and the Distinct Count for that field.

- 1 If you chose only one data field, there will be no additional rows in the upper table.
- 1 If you chose multiple data fields, there will be one row in the upper table for each field.

The Distinct Count value reflects the number of distinct values that occur in the data records for each selected data field.

Note: Both the "Field / Column" and "Distinct Count" columns can be sorted in ascending and descending order by clicking on the yellow column heading in the upper table.

Lower Table

The lower table of the Results tab displays a list of the Distinct Values for each data field as you click on each row in the upper table.

Note: Each of the three columns can be sorted in ascending and descending order by clicking on the yellow column heading in the lower table.

Pie Charts

Along with the tables describing the results, a pie chart is also displayed on the Results tab. This pie chart provides a visual representation of the statistics as you click each row in the upper table.

Note: If the Distinct Values results display as very narrow sections in the pie chart, you may want to maximize the size of the Data Inspector window for better viewing. The splitter bar between the tables and the pie chart can also be adjusted left and right to re-size each side of the Results tab.

Clear the Analysis

After running one analysis, if you want to "start over" and run another analysis, follow these steps:

1. On the Results tab, click on the yellow "Field/Column" column heading to sort the upper table by the field names.
2. Once sorted, the `_Record_Count` row should be at the top of the list. Click once on that row so no pie chart displays in the right side of the Results tab.
3. Click on the Clear Results button in the button bar, and verify that you want to close the current analysis without saving it.
 - o If you want to save the current analysis, click Yes when prompted, and save the current analysis.
 - o If you do not want to save the current analysis, click No when prompted.

Save the Analysis

After designing and running a Distinct Values analysis in the Data Inspector, you may save the analysis by following these steps:

1. Click on File in the menu bar.
2. In the Windows common dialog box, navigate to the drive and folder where you want to save the analysis.
3. Enter a name for the analysis script that will be saved. Do not overwrite the default file extension of .dp.xml.
4. Click the Save button to save the analysis script.

To open a saved analysis, follow these steps:

1. Click on File in the menu bar.
2. In the Windows common dialog box, navigate to the drive and folder where your analysis script is saved.
3. Select the desired script file with a .dp.xml file extension.