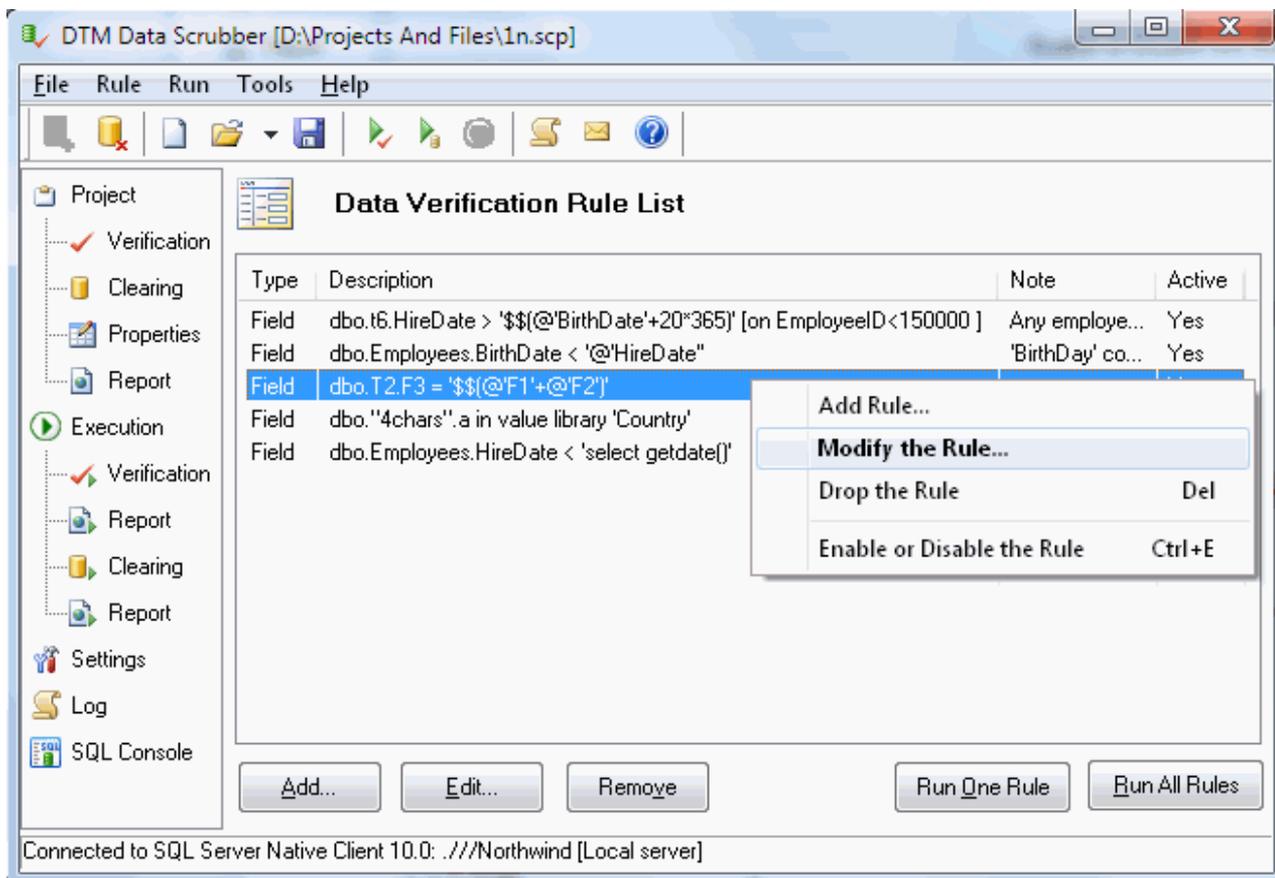


About the database validation and cleaning software

DTM Data Scrubber is a set of intelligent tools for data verification (audit) and scrubbing (cleaning). Depending on user-defined rules and data properties, the program either creates a report about the actual state of affairs or performs database data correction.

The program supports two sets of rules - for [verification](#) and for [cleaning](#). The rules of the first type describe what the correct data must look like. The rules of the second type describe the types of problems that can happen to data values. Together with this type you can specify the action, the program should perform for problem correction. There can be several rules of this type for each field and the program will perform the [action](#) corresponding to the detected conditions.

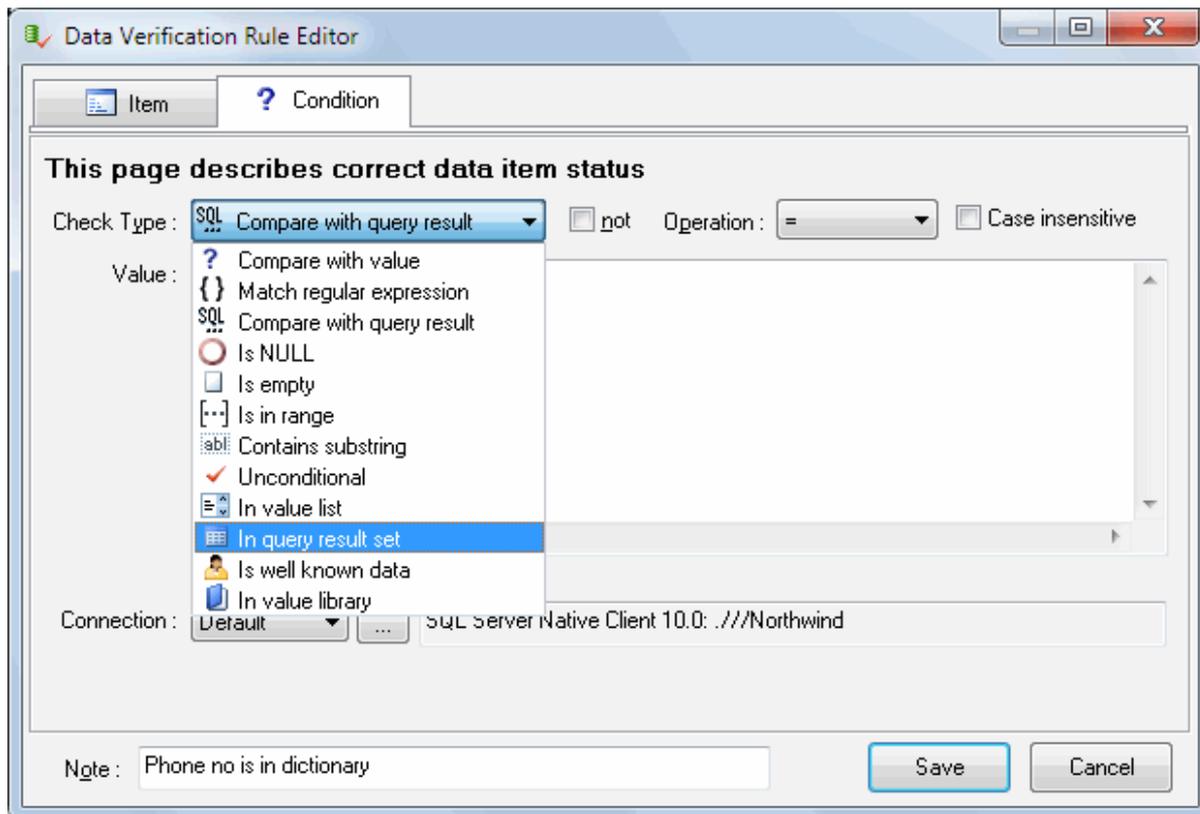
The program generates a [report](#) based on the results of verification or scrubbing. Current version of the tool supports HTML, Excel or tab-delimited text report formats.



DTM Data Scrubber is a Windows application, known to be compatible with the following operating systems: Windows XP, Windows 2003 and newer Server family and Windows Vista, 7, 8/10 (desktop). The 64-bit edition is also available.

Verification Rules

The data verification rules describe what the correct data must look like. The program will add all field values of the selected table that don't match the criterion set in the rule, to the report.



"Not" modifier reverse the condition. For example "is null" with "not" option means "is not null".

"Ignore case sensitivity" options means the program should compare string values without case. "MAR" equals to "mar" and "Mar" as example.

You are allowed to create rule for View as well as for table. Please switch on related [option](#) to access list of available views.

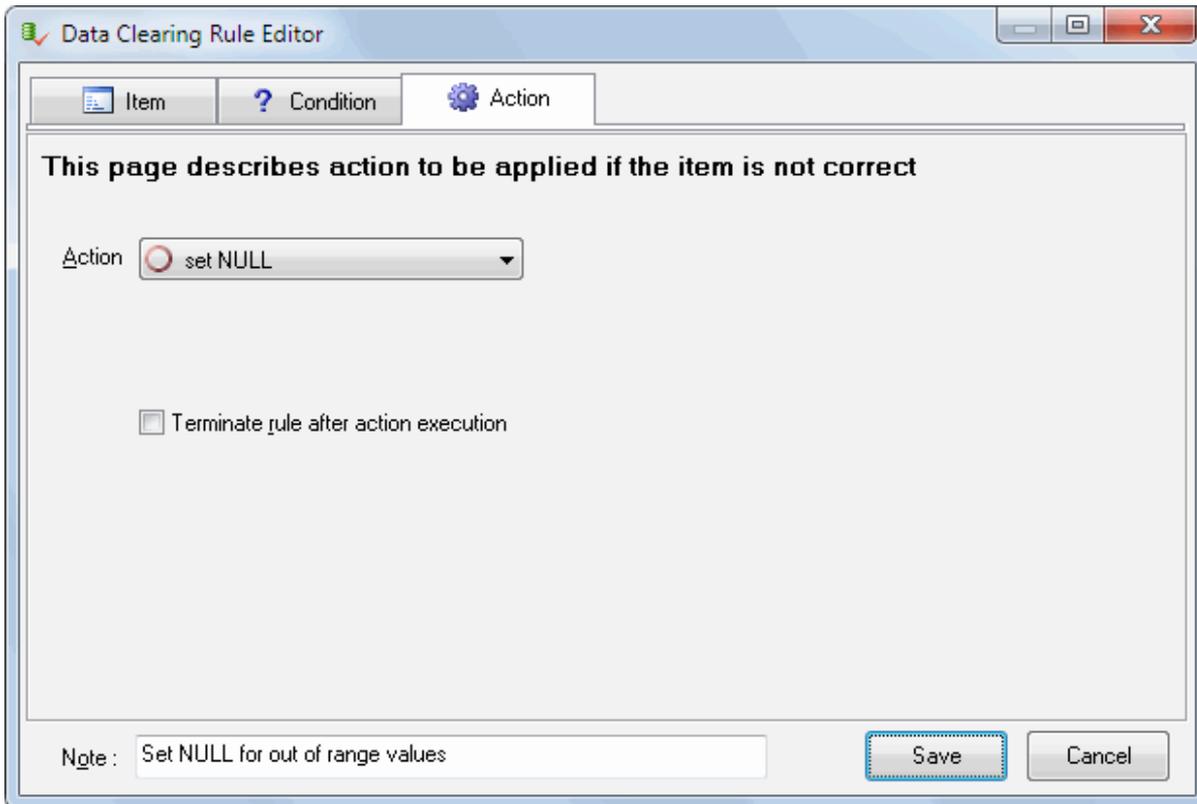
See also: unique [key selection](#).

Scrubber Rules

The Data Scrubbing rules describe the types of problems that can happen to data values. You can specify the [action](#) the program should perform for problem correction.

The [program](#) will perform the action for all records where field values satisfy the specified criterion. If the rule is marked as "terminate" and its conditions are fulfilled, this field of the current record is not checked anymore.

For example, you want to replace all empty field values with the figure 10. In this case, you should select "Is this value empty" for checking, and specify replace as an action, with 10 as a parameter. For most popular DBMS you can use built-in functions as parameters and in expressions.



"Not" modifier reverse the condition. For example "is null" with "not" option means "is not null".

"Ignore case sensitivity" options means the program should compare string values without case. "MAR" equals to "mar" and "Mar" as example.



Rule Editor

The rule editor has three tabs: [Item](#), [Condition](#) and [Action](#). The last one is applicable for data cleaning rules only.

Rule Editor, Item

This editor page helps users to specify data item to be verified. Currently, the program supports fields and results of the SQL statement execution.

Item Condition Action

This page describes item to be verified or cleaned

Single Table's Column

Table Name : Column :

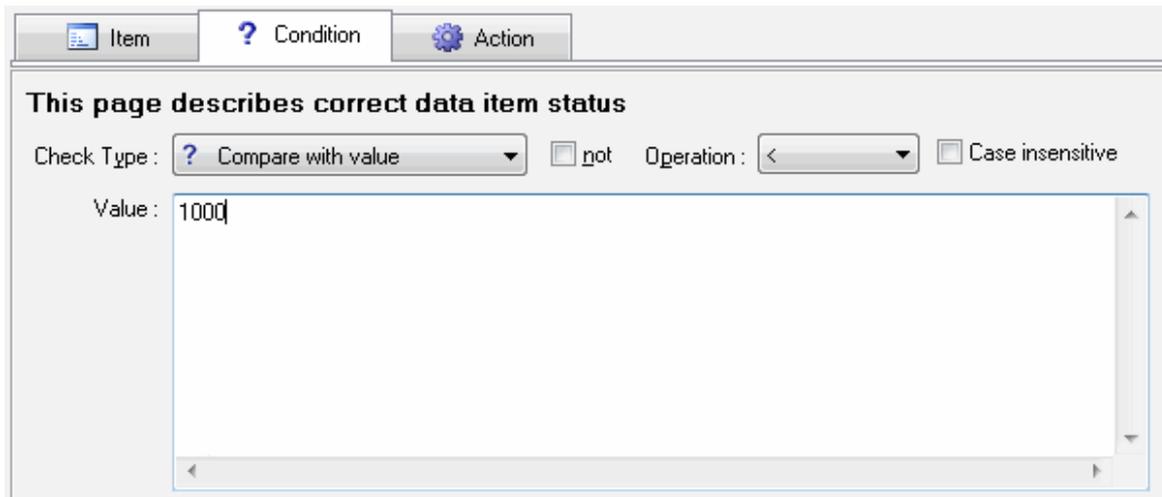
Unique key : ...

Where : optional, empty means 'all rows'

Result of the following query execution :

Rule Editor, Condition

This editor page helps users to define valid data item state. The page allows users to define [check type](#) and enter required check properties.



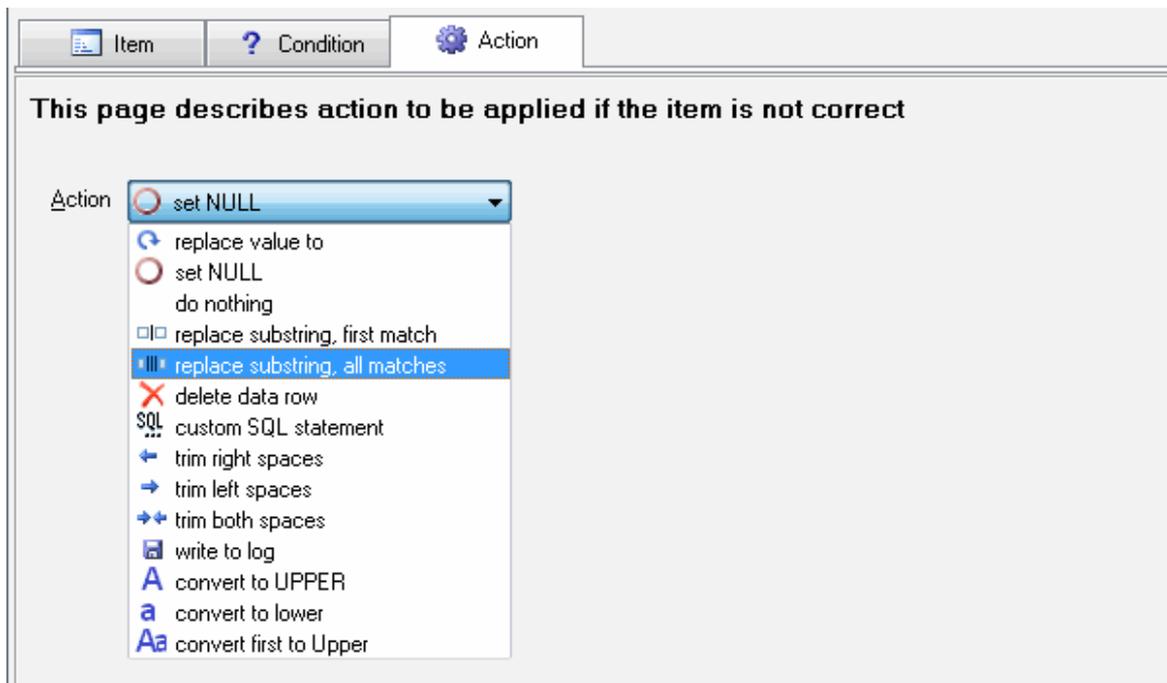
Item Condition Action

This page describes correct data item status

Check Type : ? Compare with value not Operation : < Case insensitive

Value : 1000

This editor page helps users to define action for invalid data items. It will be applied to any wrong item of data.



Quick Start: [how to connect?](#)

There are five ways to connect to a database:

1. [Direct connection](#)
2. Connection to [desktop files](#)
3. [Data source](#) with ODBC, IDAPI or Oracle Call Interface (OCI)
4. [DSN File](#) connection
5. [OLE DB](#) connection

In all modes the "Test" and "Information" buttons, as well as tools for working with connection [profiles](#) are available. "Test" button allows you to check information you entered and/or data source (or alias) configuration.

See also:

- Troubleshooting [guide](#)
- Connection [information](#)
- Connection [profiles](#)

Connection Quick Start Guide

Database	How to connect
MS SQL Server	Enter or select server name at the direct connection panel
Local SQL Server Express	Enter .\SQLEXPRESS as server name at the direct connection panel
Oracle	1) Switch to data sources mode 2) select OCI as "Interface" 3) select your service name from data source drop-down menu
DB2	Use direct connection panel or Use predefined ODBC_DSN for custom connection settings
MySQL	Install ODBC driver for MySQL from www.mysql.org Use direct connection panel or Use predefined ODBC_DSN for custom connection settings
PostgreSQL	Use direct connection panel or Use predefined ODBC_DSN for custom connection settings
Interbase/Firebird	Install ODBC driver Use direct connection panel or Use predefined ODBC_DSN for custom connection settings
Microsoft Access	1) Switch to " Desktop File " panel 2) Select "Access" as file type, enter or select file name
Microsoft Excel	1) Switch to " Desktop File " panel 2) Select "Excel" as file type, enter or select file name
<i>Another database</i>	1) Install ODBC driver for your database system 2) Create ODBC data source name using Windows ODBC Administrator 3) Switch to data sources mode 4) select your data source from drop down menu

Direct Connection

The direct connection method allows you to connect to most popular databases ([MS SQL Server](#), [Oracle](#), [Interbase/Firebird](#), [MySQL](#), [PostgreSQL](#) and [DB2](#)).

Enter the server name and the database name, if required. The user name and password are optional. Their necessity depends on the settings of your database. The owner name (schema) is optional too. The list of visible database objects depends on the choice of the owner. If the owner is empty, you will access all objects. There is important that schema/owner name is case sensitive.

If you do not find the required database type in the list or cannot connect directly, use a connection through the predefined data source. If DBMS is in the list, but unavailable, it means that either the required ODBC driver is not installed or it is not configured properly. During its use, the program stores the entered values of server names, users and owners. You can select a value from the stored list using the corresponding combo box. For some DBMS types (MS SQL, for example), the program can fill the list of available databases. Use the button with two arrows for this purpose.

Direct Database Connections to most popular databases

Microsoft SQL Server

Oracle Database

MySQL

Interbase/Firebird

IBM DB2

PostgreSQL

Note: If you do not find the required database type in the list or cannot connect directly, use a connection through the predefined Data Source.

Server: DTM-XP\SQLEXPRESS

Windows Authentication (trusted connection)

User (login): Admin

Password:

Database: Northwind

Owner/schema: dbo

Note: Express @XP

DBMS-specific connection options

Microsoft SQL Server

- "(local)", empty or "." server name means local server
- use <server name>\<instance name> syntax to identify instance. Example: .\SQLEXPRESS means SQL Express at the local system

Oracle

Use connect string for the Oracle Server that you want to access as a Server name.

Important: it is strongly recommended to use native Oracle Call Interface ([OCI](#)) instead of direct connection.

Interbase and Firebird

Examples:

- Server: **localhost** and Database **c:\interbase\myDb.fdb** - connect to specified DB on local system.
- Server: **172.17.2.10/3051** and Database **/usr/local/db/myDb.fdb** - connect to specified server with alternate port 3051 on remote system 172.17.2.10

MySQL

- Use **localhost** for local MySQL
- `example.com;port=3306` means MySQL at `example.com` on 3306 port

DB2

ServerName;port=5000;protocol=TCPIP as a server name means connect to `ServerName`, use 5000 port and TCP/IP protocol.

PostgreSQL

ServerName as a server name means connect to `ServerName`, use 5432 port and TCP/IP protocol. Database name is required. `localhost` as a server name is acceptable. To specify custom port you should add `";port=NNNN"` string to server name.

server_name_or_ip-address;port=5432;DATABASE=dbname

Desktop Files

The second way is designed for connecting to desktop data files. Select the required format and specify the file name or the directory where the data is located. Other parameters are optional.

Connections to Desktop Data File

Text file (*.txt, *.csv) Format : CustomDelimiter ANSI Delimiter : |

Microsoft Access file (*.mdb, *.accdb) use Microsoft Jet (ODBC by default)

dBase, FoxBase or FoxPro file (*.dbf)

Microsoft Excel file (*.xls, *.xlsx, *.xlsb)

Paradox file (*.db)

FoxPro database container (*.dbc)

SQLite database

Location: D:\Projects And Files\tickets.mdb Browse...

Authentication information, optional : Read Only mode

User / Login:

Password: Note: Tickets

Predefined data sources: ODBC, IDAPI, Oracle Call Interface

A connection with the use of a data source is the most universal. You can select ODBC, IDAPI or OCI (if installed) interface and the preconfigured data source name. In this case, other options are similar to those of a direct connection. The "Manage" button allows you to get access to the external configuration utility if it is available. When you want to access the tables belonging to the single database schema (or owner), you should fill the "owner" entry; otherwise, all tables will be accessed.

Connections to existing and configured data sources

Interface	<input type="text" value="ODBC"/>	<input type="checkbox"/> Manual commit
Data source	<input type="text" value="localserver"/>	<input type="button" value="Manage..."/>
User (login)	<input type="text" value="sa"/>	
Password	<input type="text" value="xxxxxx"/>	
Database	<input type="text" value="Northwind"/>	
Owner	<input type="text" value="dbo"/>	
Note	<input type="text" value="Local SQL Server"/>	

DSN File

The fourth way is using a DSN file. For this case, just select the file name with DSN definition.

Connections to existing and configured file DSN

File DSN name

Note

OLE DB connection

Use 'Configure' button to specify connection information. Password and owner fields are optional.

Connections using OLE DB providers

Connection properties Configure...

```
Provider=SQLOLEDB.1;  
Persist Security Info=False;  
User ID=sa;  
Initial Catalog=Northwind;  
Data Source=DTM-ACER2;  
Use Procedure for Prepare=1;  
Auto Translate=True;  
Packet Size=4096;
```

Password

Owner

Note

Connection Profiles

Connection profile helps you to save information about your connection (interface, data source or alias name, user name (login), password and database name, etc) and get access it by the one click.

Please fill connection properties and press "Add as new" to add a new profile. To modify the profile you should select it from the list at the top of the window, modify properties and press "Update". "Delete" button works when you select the profile to be deleted in the list.

"Save" and "Load" buttons allow you to save profiles to the disk file or load them. The "Export one" button helps to save single currently selected profile.

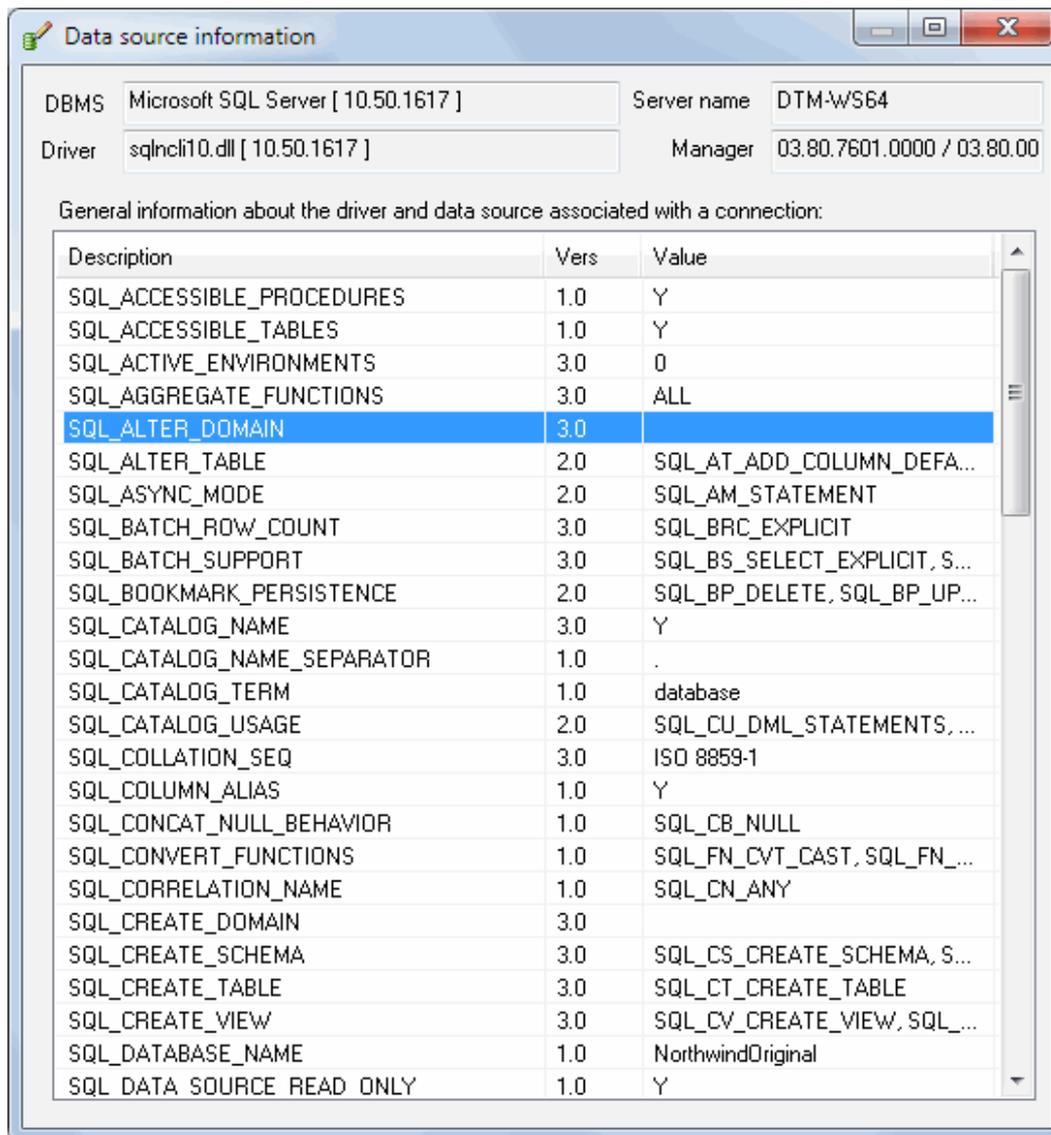
Important: all profiles are shared between all installed DTM soft products. That means once created profile can be used with any tool. At the other side if you remove the profile from the list you can't use it with DTM soft's products anymore.

Mode	Interface	Source or Server	User	Owner	Database	Note
Direct	ODBC	.			Northwind	Local server
Direct	ODBC	DTM-XP	sa		AdventureWorks	
Direct	ODBC	.		dbo	NorthwindOriginal	Local Read Only DB
DSN	ODBC	Saramdb				SaraMDB
DSN	ODBC	test_new				Test MDB
Desktop	ODBC	ACCESS				Test MDB
DSN	ODBC	test_old				Test MDB
Direct	ODBC	.		dbo		
Direct	ODBC	.		dbo		2
DSN	OCI	ORCL	OE			ORCL/OE
DSN	ODBC	ORCL_ODBC	OE	OE		ORCL/ODBC
DSN	OCI	10G	OE			Oracle 10g (VM)

Buttons: Add as New, Update profile, Remove profile, Export One..., Save..., Load...

Connection Information

The program provides detailed database, connection and driver information and properties. The "Information" button at the connect window allows you to view it.



The screenshot shows a window titled "Data source information" with the following fields:

- DBMS: Microsoft SQL Server [10.50.1617]
- Server name: DTM-WS64
- Driver: sqlncli10.dll [10.50.1617]
- Manager: 03.80.7601.0000 / 03.80.00

General information about the driver and data source associated with a connection:

Description	Vers	Value
SQL_ACCESSIBLE_PROCEDURES	1.0	Y
SQL_ACCESSIBLE_TABLES	1.0	Y
SQL_ACTIVE_ENVIRONMENTS	3.0	0
SQLAggregate_FUNCTIONS	3.0	ALL
SQL_ALTER_DOMAIN	3.0	
SQL_ALTER_TABLE	2.0	SQL_AT_ADD_COLUMN_DEFA...
SQL_ASYNC_MODE	2.0	SQL_AM_STATEMENT
SQL_BATCH_ROW_COUNT	3.0	SQL_BRC_EXPLICIT
SQL_BATCH_SUPPORT	3.0	SQL_BS_SELECT_EXPLICIT, S...
SQL_BOOKMARK_PERSISTENCE	2.0	SQL_BP_DELETE, SQL_BP_UP...
SQL_CATALOG_NAME	3.0	Y
SQL_CATALOG_NAME_SEPARATOR	1.0	.
SQL_CATALOG_TERM	1.0	database
SQL_CATALOG_USAGE	2.0	SQL_CU_DML_STATEMENTS, ...
SQL_COLLATION_SEQ	3.0	ISO 8859-1
SQL_COLUMN_ALIAS	1.0	Y
SQL_CONCAT_NULL_BEHAVIOR	1.0	SQL_CB_NULL
SQL_CONVERT_FUNCTIONS	1.0	SQL_FN_CVT_CAST, SQL_FN_...
SQL_CORRELATION_NAME	1.0	SQL_CN_ANY
SQL_CREATE_DOMAIN	3.0	
SQL_CREATE_SCHEMA	3.0	SQL_CS_CREATE_SCHEMA, S...
SQL_CREATE_TABLE	3.0	SQL_CT_CREATE_TABLE
SQL_CREATE_VIEW	3.0	SQL_CV_CREATE_VIEW, SQL_...
SQL_DATABASE_NAME	1.0	NorthwindOriginal
SQL DATA SOURCE READ ONLY	1.0	Y

Troubleshooting Guide

Problem description	Possible reason	Solutions
Required database type not present in the list at Direct Connection and Desktop Connection pages		Switch to "data source" connection mode and select data source from the list or configure new one with "Manage" button.
Required format is in the direct connection list, but not available (disabled).	ODBC driver for your database does not installed or not configured properly.	Install required driver. If it is already present in the system, please contact our support staff .
Errors during direct connection.	Compatibility problems.	Try to create data source for your database connection.
Login error for correct user name and password.	Read-only desktop data file.	Try to change file mode to 'read and write'.
I can't see relationships, defaults, etc in my Access Database.	Access interface.	Try to switch on "Use Microsoft Jet" check box at the "Desktop File" page of the Connect Window.



Types of Check-up

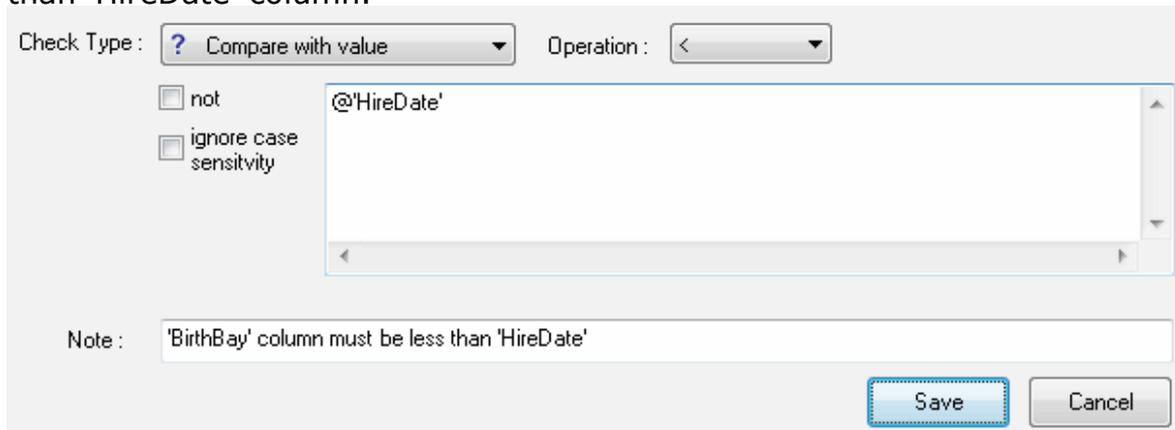
The [program](#) supports the following types of data checks:

- Compare the field value with a constant, another [field](#) or [expression](#)
- Are there values matching a regular expression
- Compare the field value with an SQL statement execution result. The program uses the first value of the first column.
- Does the value NULL
- Does the value empty
- Does the value within the range of values
- Does the value contain substring
- Does the value in the specified list. Use ',' as value separator. Non -significant spaces are not allowed.
- Does the value in the result set of SQL statement execution. The data scrubber uses the first result set only.
- Does the value has well-known format like URL, e-mail, phone number, etc.
- Find the value in the list of predefined data like countries, codes, currencies, etc. This collection called 'Value library'.
- Check that the SQL statement from [items](#) page successfully executed.

Also, there is "unconditional" check. It means the cleaning [action](#) will be executed for any data item.

Compare with another Field

@'filed name' as value allows the user to specify internal row integrity checks. In the following example the rule checks that 'BirthDate' column of Employees table is less than 'HireDate' column.



Check Type : ? Compare with value Operation : <

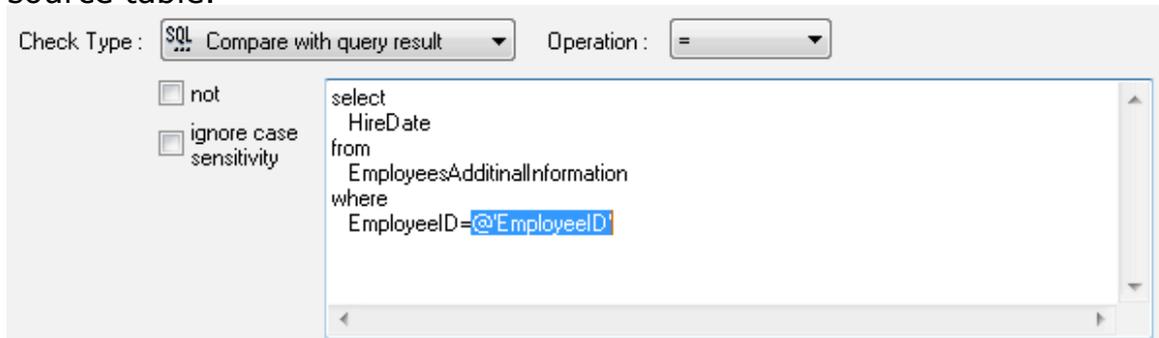
not @'HireDate'

ignore case sensitivity

Note : 'BirthDate' column must be less than 'HireDate'

Save Cancel

The reference to column can also be used in the SQL statement as shown at the following screenshot. Please note that it works slow due to rerun query for each data row of the source table.



Check Type : SQL Compare with query result Operation : =

not

ignore case sensitivity

```
select
  HireDate
from
  EmployeesAdditionalInformation
where
  EmployeeID=@'EmployeeID'
```

Expression as a value

The user can specify expression instead of immediately entered value of reference to another [field](#).

Syntax: \$\$ (expression text)

An expression can contain: constants, operations and references to field value as '@field name'.

Operations

Operations '*', '/' and unary '-' are acceptable for numeric operands only.

The '+' operation

Left operand	Right operand			
	Integer	Float	Date	Time
Integer	+	+ (float)	Add N days	Add N seconds
Float	+ (float)	+	N/A	N/A
Date	add N days	N/A	N/A	N/A
Time	add N seconds	N/A	N/A	N/A

The '-' operation

Left operand	Right operand			
	Integer	Float	Date	Time
Integer	-	- (float)	N/A	N/A
Float	- (float)	+	N/A	N/A
Date	subtract N days	N/A	N/A	N/A
Time	subtract N seconds	N/A	N/A	N/A

Notes and limitations:

- '(' and ')' can be used with any acceptable data type.
- Complex nested expressions like \$\$(\$\$(8-2)-2) are not acceptable, use '()' instead of. Example: \$\$((8-2)-2)

Example: 'HireDate' must be greater than 'BirthDate' + 18 years (265*18 days):

Table Name :	<input type="text" value="dbo.Employees"/>	Column :	<input type="text" value="HireDate"/>
Where :	<input type="text"/>	optional, empty string means "all rows"	
Check Type :	<input type="text" value="Compare with value"/>	Operation :	<input type="text" value=">"/>
<input type="checkbox"/> not	<input type="text" value="\$\$(@'BirthDate'+18*365)"/>		
<input type="checkbox"/> ignore case sensitivity			
Note :	<input type="text" value="Any employee must be at least 18 y.o."/>		
		<input type="button" value="Save"/>	<input type="button" value="Cancel"/>

A regular expression is a special template type allowing you to create effective [checks](#) and [substring changes](#) in data. An expression consists of usual characters and meta symbols.

Most popular meta symbols and language clauses are:

Item	Match with
^	the beginning of line
\$	the end of line
.	any character
[]	characters in set
[^]	characters not in set
	previous or next pattern
?	previous pattern 0 or 1 times
+	previous pattern 1 or more times
*	previous pattern 0 or more times
()	group pattern and remember to \n variable

If part of the expression is enclosed in brackets, after the matching is over, the substring value gets into a variable of the type \n, where n is the number of the expression from the beginning of the string. These variables can be used only in replace [actions](#).

Useful regular expressions

Expression	Description
^[0-9A-Fa-f]+\$	Hexadecimal number
^\(\(\d{3}\)\) ?\d{3}[-]\d{4}\$	Phone number (NNN) NNN-NNNN
^\^+([1-9] \(\(\d{3}\)\) ?\d{3}[-]\d{4}\$	International phone number +N (NNN) NNN-NNNN
^http\:\/\/[a-zA-Z0-9\-\.\.]+\.[a-zA-Z]{2,3}(/\S*)?\$	URL
^[0-9a-zA-Z_\.\.]+@([0-9a-zA-Z_\.\.]+?\.\.){1,2}[a-zA-Z]{2,4}\$	E-mail

Types of Actions

The [program](#) supports the following types of actions:

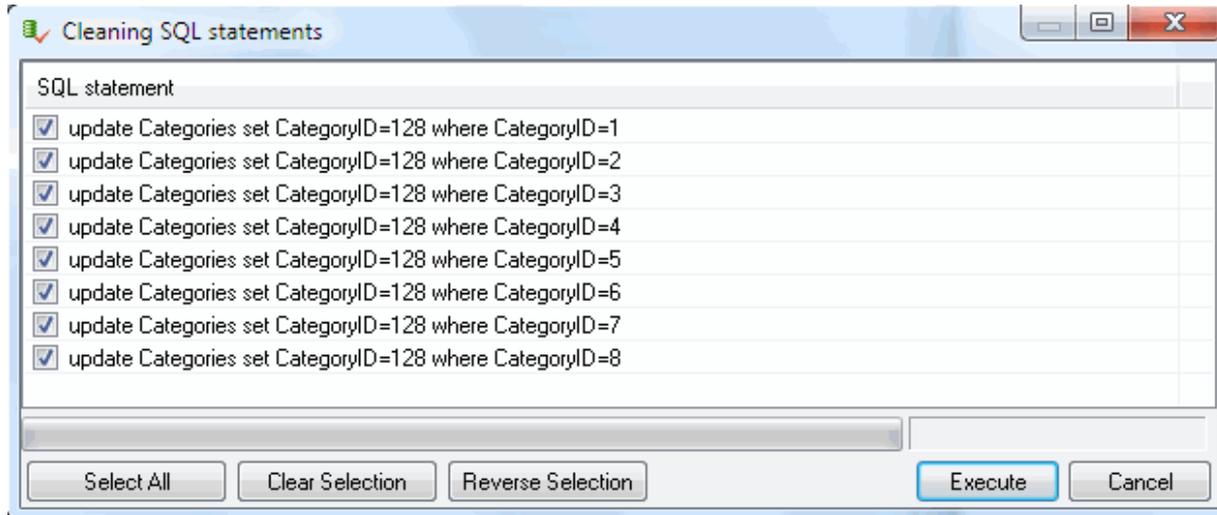
- Changing the value to a constant or an expression (only for match with a [regular expression](#)).
- Setting a value for a NULL field.
- Do nothing. Usually used for terminate rule or unconditional trims.
- Replace substring (first match), works with "contains substring" check only.
- Replace substring (all matches), works with "contains substring" check only.
- Remove the row.
- Custom SQL statement. Use %WHERE% macro to identify current data row.
- Trim spaces: left, right or both.
- Write message to product's log file.
- String conversion: make all letters upper [AAAAA]
- String conversion: make all letters lower [aaaaa]
- String conversion: make the first letter of all words upper and next lower [Aaaa Aaa]

%VALUE% macro can be used in the action description and means original data item value. Example for "Custom SQL statement" action type:

```
update OrderDetails set Discount=%VALUE%*2 %WHERE%
```

Data Cleaning Script

In the data scrubbing mode, the program creates a set of SQL statements for database correction. The user can select the necessary statements or forbid the unnecessary ones in a special dialog box.



Important: [console mode](#) of the tool does not show this window and executes all cleaning SQL statement unconditional.

Execution log (console) is a list of the executed items of the project. It allows you to view execution process and results of the project execution. There are two logs present (for verification and for data scrubbing mode).



Data Verification Rules Execution

Database Object	Rows	Rules Checked	Not Matched
Table 'dbo.Employees', 10 rows	10	10	4

This option helps the user to view execution report.

Please keep in mind that you can use it if specify corresponding report file at the [project properties](#) page.

 **Data Verification Report**

Execution Report for Verification

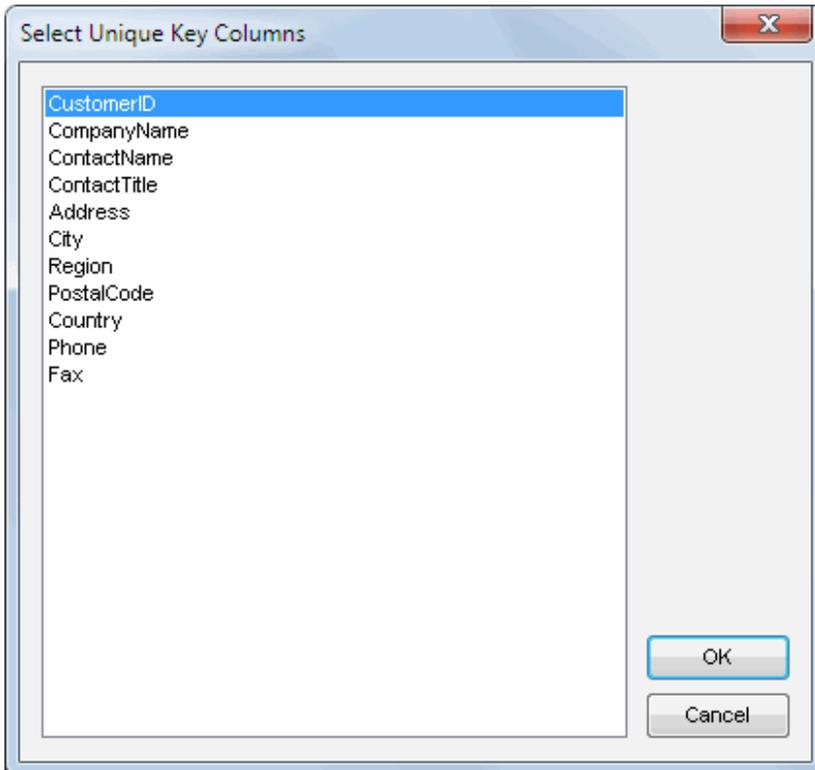
Property	Value
DTM Data Scrubber	1.08.02
Report Created	18-Jun-2011 09:39:21
Connection	SQL Server Native Client 10.0: ///Northwind [Local server]

Table 'dbo.Employees'

Row	Key: EmployeeID	Column Name	Wrong Value	Applied Rule
2	6502	BirthDate	1964-08-18 15:06:34.000	'BirthDay' column must be less than 'HireDate' (1955-06-12 06:55:37.000)
3	6503	BirthDate	2008-08-29 13:36:19.000	'BirthDay' column must be less than 'HireDate' (1974-11-06 16:09:00.000)
4	6504	BirthDate	NULL	'BirthDay' column must be less than 'HireDate'
10	6510	BirthDate	1981-10-08 04:52:00.000	'BirthDay' column must be less than 'HireDate' (1950-09-15 21:28:01.000)

Unique key selection

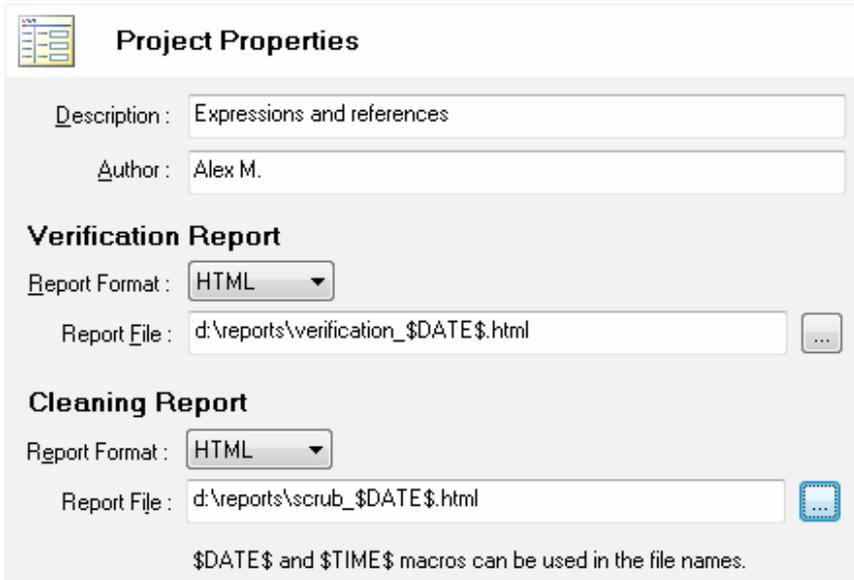
To identify the wrong row your table must have primary or alternate unique key. The program uses primary key automatically by default. However, users allowed to change unique key or assign it manually.



Project Properties

There are four important project properties:

1. Project description (optional)
2. Project author (optional)
3. The format of verification report (HTML, Excel or tab-delimited text)
4. The report file name for verification process
5. The format of cleaning report (HTML, Excel or tab-delimited text)
6. Report file name for cleaning process



Project Properties

Description : Expressions and references

Author : Alex M.

Verification Report

Report Format : HTML

Report File : d:\reports\verification_$\\$DATE\\$\$.html

Cleaning Report

Report Format : HTML

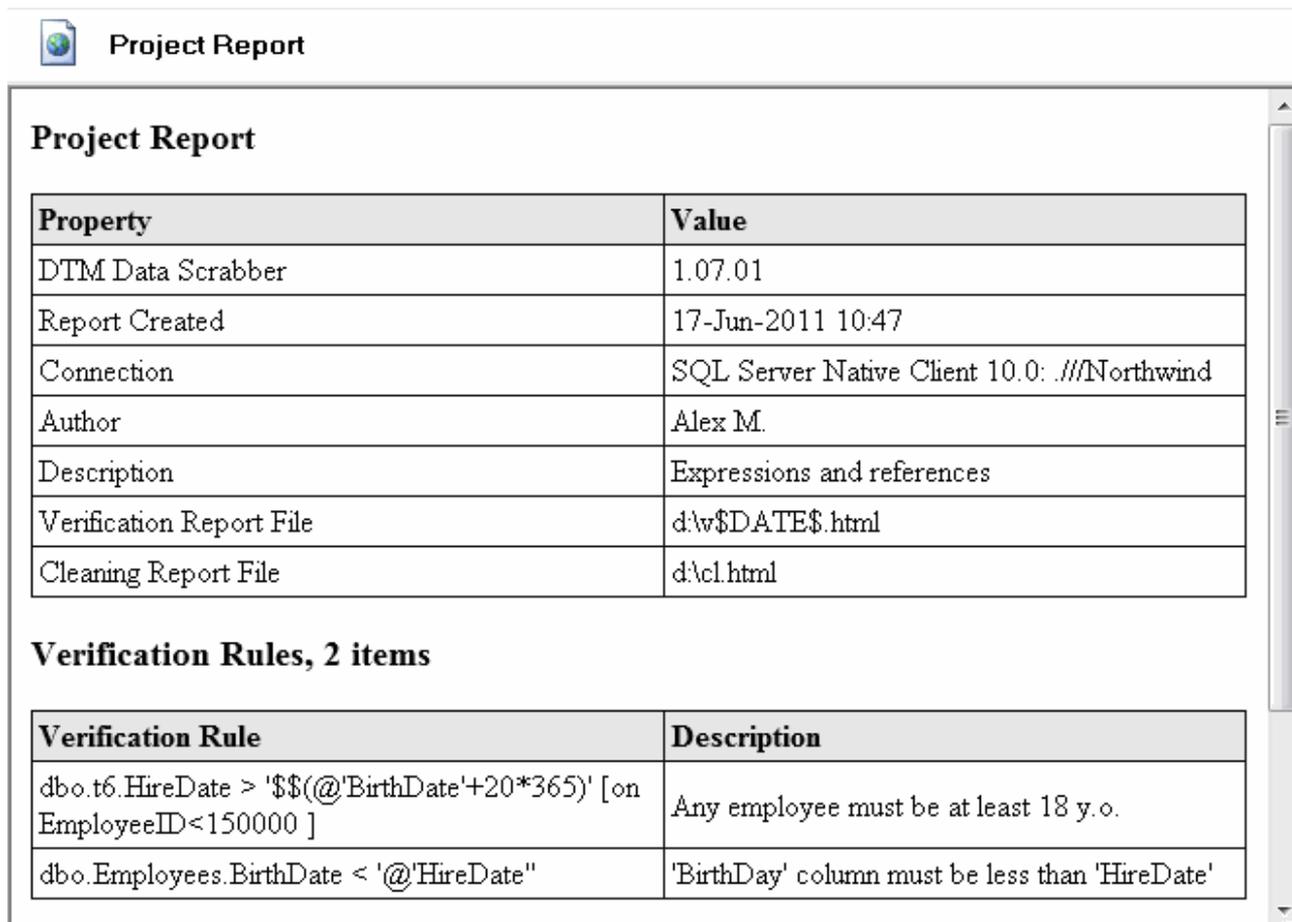
Report File : d:\reports\scrub_$\\$DATE\\$\$.html

$\\$DATE\\$\$ and $\\$TIME\\$\$ macros can be used in the file names.

Notes:

- Please refer to project [report](#) to view all properties of the project including definitions of the rules.
- You can use $\\$DATE\\$\$ and $\%TIME\\$\$ macros in the report file name.

The project report is a good way to review all rules. It provides summary information about the project as well as overview for each rule of the project.



The screenshot shows a window titled "Project Report" with a globe icon. The window content is divided into two sections. The first section, titled "Project Report", contains a table with two columns: "Property" and "Value". The second section, titled "Verification Rules, 2 items", contains a table with two columns: "Verification Rule" and "Description".

Property	Value
DTM Data Scrubber	1.07.01
Report Created	17-Jun-2011 10:47
Connection	SQL Server Native Client 10.0: ///Northwind
Author	Alex M.
Description	Expressions and references
Verification Report File	d:\w\$DATE\$.html
Cleaning Report File	d:\cl.html

Verification Rule	Description
dbo.t6.HireDate > '\$\$(@'BirthDate'+20*365)' [on EmployeeID<150000]	Any employee must be at least 18 y.o.
dbo.Employees.BirthDate < '@'HireDate"	'BirthDay' column must be less than 'HireDate'

See also: project [properties](#).



Command line options

DTM Data Scrubber supports following command line switches:

- r** - if project configured, run clearing at program startup
- v** - if project configured, run verification at program startup
- q** - quit application after the project execution
- c** - [console mode](#).

Also, you can use project name as a command line parameter. Please note that you should use quotation for path or filename with spaces:

"c:\Program Files\DTM Data Scrubber\scr.exe" "D:\Projects And Files\test2testscr.scp"



Console Mode

The console mode is a mode when DTM Data Scrubber doesn't open any dialogs and doesn't need any interference from the user. A project for the console mode must be prepared and tested beforehand. This mode enables you to integrate the product with the Windows task scheduler as well as to execute projects prepared beforehand according to the schedule.

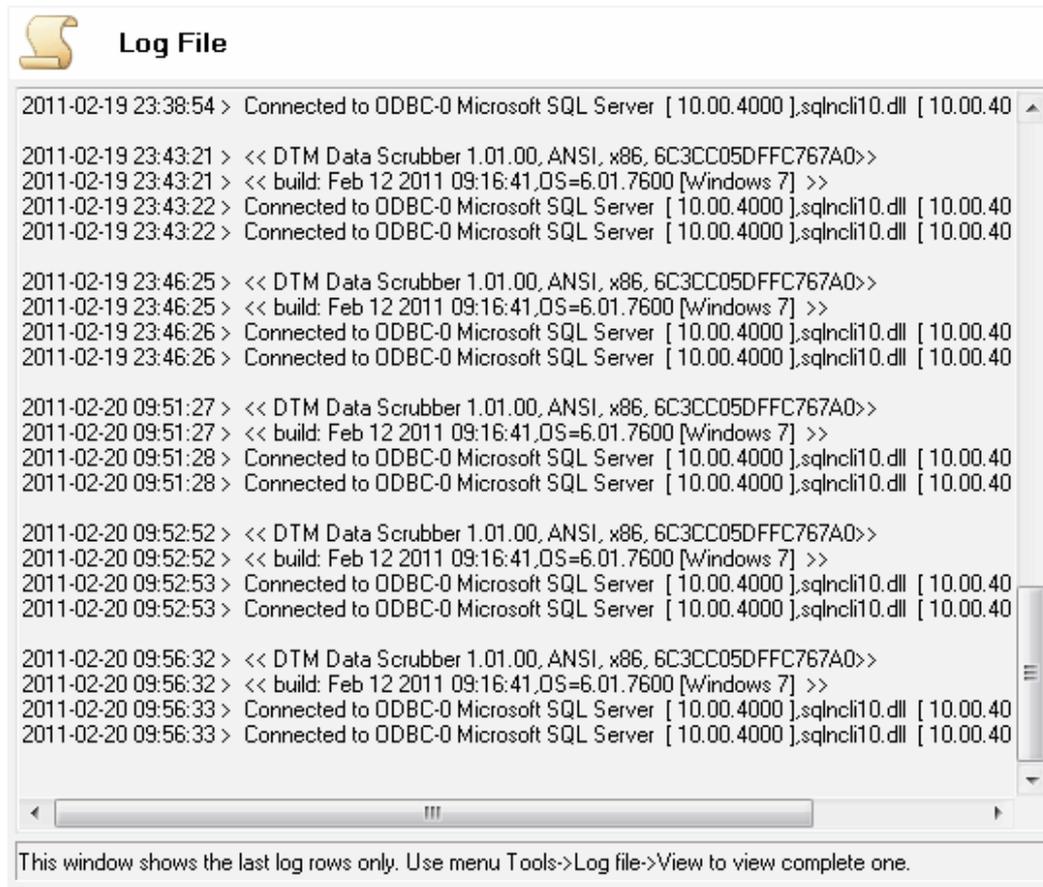
For example, in order to execute a data verification process today at 11.00, use the following command line:

at 11:00 "scr.exe -c -v"

See also: [command line](#) switches.

Built-in log viewer

This panel shows last rows of the [program log](#).
To view complete one please use "Tools->Log" menu item or Ctrl+L [hotkey](#).



The screenshot shows a window titled "Log File" with a scrollable list of log entries. The entries are as follows:

```
2011-02-19 23:38:54 > Connected to ODBC-0 Microsoft SQL Server [ 10.00.4000 ],sqlncli10.dll [ 10.00.40
2011-02-19 23:43:21 > << DTM Data Scrubber 1.01.00, ANSI, x86, 6C3CC05DFFC767A0>>
2011-02-19 23:43:21 > << build: Feb 12 2011 09:16:41,DS=6.01.7600 [Windows 7] >>
2011-02-19 23:43:22 > Connected to ODBC-0 Microsoft SQL Server [ 10.00.4000 ],sqlncli10.dll [ 10.00.40
2011-02-19 23:43:22 > Connected to ODBC-0 Microsoft SQL Server [ 10.00.4000 ],sqlncli10.dll [ 10.00.40

2011-02-19 23:46:25 > << DTM Data Scrubber 1.01.00, ANSI, x86, 6C3CC05DFFC767A0>>
2011-02-19 23:46:25 > << build: Feb 12 2011 09:16:41,DS=6.01.7600 [Windows 7] >>
2011-02-19 23:46:26 > Connected to ODBC-0 Microsoft SQL Server [ 10.00.4000 ],sqlncli10.dll [ 10.00.40
2011-02-19 23:46:26 > Connected to ODBC-0 Microsoft SQL Server [ 10.00.4000 ],sqlncli10.dll [ 10.00.40

2011-02-20 09:51:27 > << DTM Data Scrubber 1.01.00, ANSI, x86, 6C3CC05DFFC767A0>>
2011-02-20 09:51:27 > << build: Feb 12 2011 09:16:41,DS=6.01.7600 [Windows 7] >>
2011-02-20 09:51:28 > Connected to ODBC-0 Microsoft SQL Server [ 10.00.4000 ],sqlncli10.dll [ 10.00.40
2011-02-20 09:51:28 > Connected to ODBC-0 Microsoft SQL Server [ 10.00.4000 ],sqlncli10.dll [ 10.00.40

2011-02-20 09:52:52 > << DTM Data Scrubber 1.01.00, ANSI, x86, 6C3CC05DFFC767A0>>
2011-02-20 09:52:52 > << build: Feb 12 2011 09:16:41,DS=6.01.7600 [Windows 7] >>
2011-02-20 09:52:53 > Connected to ODBC-0 Microsoft SQL Server [ 10.00.4000 ],sqlncli10.dll [ 10.00.40
2011-02-20 09:52:53 > Connected to ODBC-0 Microsoft SQL Server [ 10.00.4000 ],sqlncli10.dll [ 10.00.40

2011-02-20 09:56:32 > << DTM Data Scrubber 1.01.00, ANSI, x86, 6C3CC05DFFC767A0>>
2011-02-20 09:56:32 > << build: Feb 12 2011 09:16:41,DS=6.01.7600 [Windows 7] >>
2011-02-20 09:56:33 > Connected to ODBC-0 Microsoft SQL Server [ 10.00.4000 ],sqlncli10.dll [ 10.00.40
2011-02-20 09:56:33 > Connected to ODBC-0 Microsoft SQL Server [ 10.00.4000 ],sqlncli10.dll [ 10.00.40
```

At the bottom of the window, there is a message: "This window shows the last log rows only. Use menu Tools->Log file->View to view complete one."

Q: How can I replace substring AAA in field value to BBB?

A: Create rule using "find substring" condition and "replace substring" as action. Use AAA as a condition value and BBB as a replace string.

A: Create rule using "match regular expression" condition and "replace value to" action. Use `(.*)AAA(.*)` as a condition value and `\1BBB\2` as a replace string.

Q: How to check that column is equal to the sum of other columns, i.e. $F3 = F1 + F2$?

A: Create rule using "compare with value" check. Use '=' as operation and `$$(@'F1'+@'F2')` expression as value.

Q: How to check that 'Sex' column has 'M' and 'F' values only?

A: Create rule using "in list" check. Use F,M as list. Switch "ignore case" check box is necessary.

Q: How to check that 'HireDate' column less or equal than current date?

A: Create rule using "compare with query result" check and "<=" as operation. Use "select getdate()" as a query.

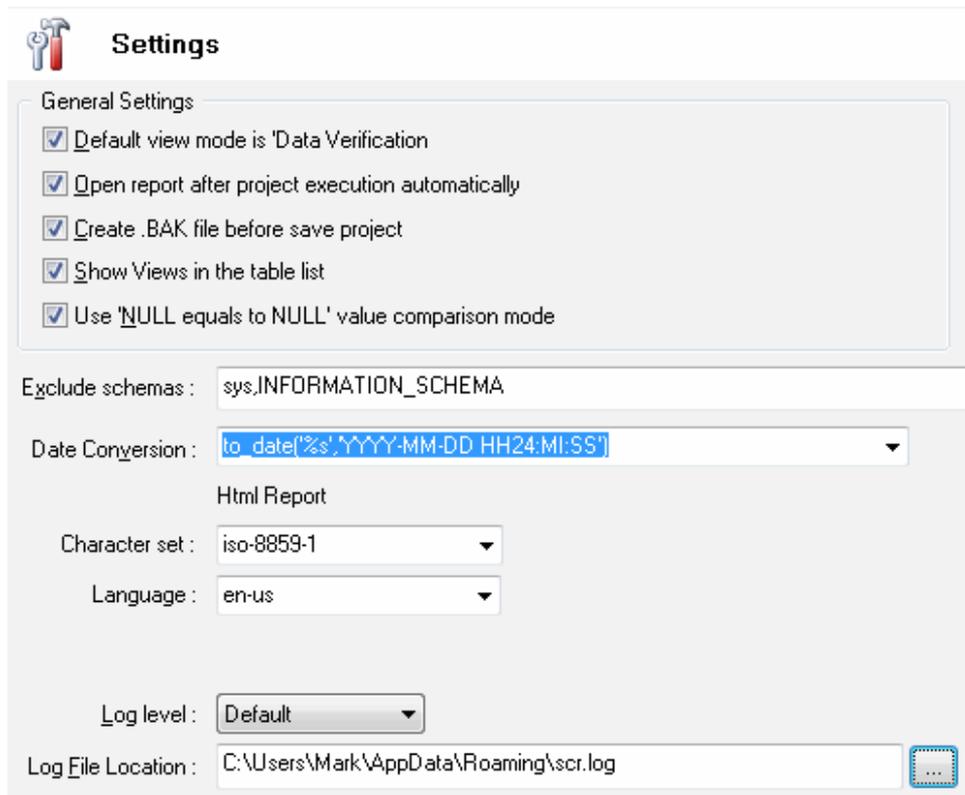
It is for SQL Server. Please change built-in function name if necessary.

Q: I have Discount column. How to divide in half all discounts greater 0.10?

A: Create rule using "compare with value" check, 0.10 as value and ">" as operation. Use "replace value to" action with `"%VALUE%/2"` value.

Product Settings

The [program](#) has a few settings and options to make your work more comfortable.



Settings

General Settings

- Default view mode is 'Data Verification'
- Open report after project execution automatically
- Create .BAK file before save project
- Show Views in the table list
- Use 'NULL equals to NULL' value comparison mode

Exclude schemas : sys,INFORMATION_SCHEMA

Date Conversion : to_date('%s', 'YYYY-MM-DD HH24:MI:SS')

Html Report

Character set : iso-8859-1

Language : en-us

Log level : Default

Log File Location : C:\Users\Mark\AppData\Roaming\scr.log

Hot Keys

Hotkey	Function or Option
Ctrl+E	Enable or disable current rule
Ctrl+F	Run the verification (audit) rules
Ctrl+L	View log file
Ctrl+N	Create new empty project file
Ctrl+O	Load Project file from the disk
Ctrl+Q	Quit the program
Ctrl+R	Run the cleaning rules
Ctrl+S	Save Project file to the disk
F1	Open help
F2	Open the connect window
F3	Close the database connection (disconnect)

The tool has a special window where you can specify and execute any SQL statements. You can copy the results of executing a statement onto the clipboard or export it into various formats like text, SQL, HTML, XML or Microsoft Excel. Placing the mouse cursor over the column header will show the type of data stored in this field.

Important: the SQL console does not show any warning before data deletion or modifying.

Note: SQL console shows only begins of large strings. Typically you can view up to 512 first symbols.

The screenshot shows the SQL Console interface. At the top, there are three buttons: "Run", "Load...", and "Export...". Below the buttons is a text area containing the SQL query: `select * from Customers`. Below the query area, there is a status bar showing "Fetching time: 0.266 sec" and "Go to 1 record". The main area displays a table with the following data:

	CustomerID	CompanyName	ContactName	ContactTitle	Address
1	ALFKI	Alfreds Futterkiste	Maria Anders	Sales Representative	Obere...
2	ANATR	Ana Trujillo Emparedados y helados	Ana Trujillo	Owner	Avda. c...
3	ANTON	Antonio Moreno Taqueria	Antonio Moreno	Owner	Matad...
4	AROUT	Around the Horn	Thomas Hardy	Sales Representative	120 Ha...
5	BERGS	Berglunds snabbkop	Christina Berglund	Order Administrator	Berguv...
6	BLAUS	Blauer See Delikatessen	Hanna Moos	Sales Representative	Forster...

Menu item "Load" allows you to read SQL script from the external file.

There is a picture of local menu accessed by the right click inside the results window.

The screenshot shows a close-up of the results window. A context menu is open over the table, with the following options: "Copy selected" and "Export...". The table data is partially visible:

	CustomerID	CompanyName
1	ALFKI	Alfreds Futterkiste
2	ANATR	Ana Trujillo Emparedados y helados
3	ANTON	Antonio Moreno Taqueria
4	AROUT	Around the Horn



License Agreement (EULA)

This License Agreement covers all existing versions of DTM Data Scrubber (Software) and technical support service (Service). This License Agreement is a legal agreement between the end-user (Licensee) and DTM soft (Licensor).

CAREFULLY READ THE TERMS AND CONDITIONS OF THIS AGREEMENT PRIOR TO USING THIS PRODUCT. USE OF ANY PORTION OF THIS PACKAGE INDICATES YOUR AGREEMENT TO THE FOLLOWING TERMS AND CONDITIONS. IF YOU DO NOT AGREE WITH SUCH TERMS AND CONDITIONS, DO NOT INSTALL THE SOFTWARE.

General Information

1. Licensor is exclusive owner of all DTM Data Scrubber copyrights. DTM Data Scrubber is protected by copyright laws and international copyright treaties.
2. Demo version. Anyone may install and use demo version of DTM Data Scrubber for evaluation and testing purposes free of charge.
3. The product is licensed, not sold. I.e. Licensor grants to Licensee non-exclusive, perpetual, royalty-free right and license to install, configure, execute and otherwise productively use a copy of the Software for the commercial or non-commercial purposes, including internal business purposes. Licensee may install and use each licensed copy of the Software on a single computer. The primary user of the computer on which DTM Data Scrubber is installed may make a second copy for his or her exclusive use on a portable computer.
4. Licensee may not reverse engineer, modify, translate, decompile, or disassemble DTM Data Scrubber. The Software is licensed as a single product. Its component parts may not be separated for use on more than one computer.
5. Licensee may not rent, lease, or lend the Software. Also, Licensee may not resell, or otherwise transfer for value, the Software.
6. Without prejudice to any other rights, Licensor may terminate this License Agreement if Licensee fail to comply with the terms and conditions of this Agreement. In such event, Licensee must destroy all copies of the Software with all of its component parts.
7. Licensee may permanently transfer all of rights under this license, provided Licensee retain no copies, Licensee transfer all of DTM Data Scrubber (including all component parts), and the recipient agrees to the terms of this license.
8. DTM Data Scrubber IS DISTRIBUTED "AS IS". NO WARRANTY OF ANY KIND IS EXPRESSED OR IMPLIED. LICENSEE USE DTM Data Scrubber AT YOUR OWN RISK. IN NO EVENT SHALL EITHER PARTY BE LIABLE FOR DATA LOSS, DAMAGES, LOSS OF PROFITS OR ANY OTHER KIND OF LOSS WHILE USING OR MISUSING THIS SOFTWARE.

Delivery

Licensor delivers the Software electronically over Internet. The delivery includes installation/activation key, software, documentation* and additional materials with installation program*. Licensor confirms that the delivery contains no illicit code or third party code.

The customers in North America allowed to request physical delivery on CD for extra fee.

* - the multiplatform edition of the software (if applicable) has online documentation only and requires no installation program.

Licensee Data

Licensor understands and acknowledges that Licensee may manage, modify, summarize, maintain, create derivative works of, and update pre-existing data and information, and generate, manage, modify, summarize, maintain, create derivative works of, and update additional data and information using the Software. Licensor acknowledges and agrees that all rights in any work product created by Licensee shall be solely owned by Licensee. Licensor has no access to mentioned work product without grant by Licensee. Moreover,

Licensor shall not seek access to Licensee personal data.

Functionality

Licensor may remove or change any supplemental or non critical functionality of the Software without any notifications. Licensor may change product documentation and project file format. If Licensor deletes any key feature or functionality from the Software equal solution should be provided to Licensee without additional fee.

Third Party Software and Intellectual Property

Licensor hereby confirms that the Software contains no third party components including Open Source code.

Support and Upgrades

During one year after ordering any license of the Software except "Site" and "World" licenses, Licensee are entitled to free technical services and support for DTM Data Scrubber which is provided by Licensor. During this period, e-mail support is unlimited and includes technical and support questions. Also, during one year, Licensee may access to free updates to DTM Data Scrubber when and as DTM soft publishes them on www.sqledit.com. After end of the described period Licensee may continue to use the software product in accordance with the terms of this Agreement except free support and upgrades. After end of the free support and updates period (one year), Licensee may purchase annual Upgrade and Support subscription. If Licensee has a few licenses, Licensee will access to free upgrade and support period and will use subscriptions independently.

There are two support service levels: Standard and Premium. By default, the Software includes Standard level of the Service. The Premium should be ordered separately. The Premium service offers reduced response time and high priority for support requests. The technical support response time for Standard level is between 2 and 12 hours except Jun, July and August with 2 to 24 hours range. It is available 5 days per week (Monday to Friday). For the Premium level it is 2 to 6 hours 7 days per week.

Export Compliance

Licensor hereby confirms that the Software requires no export controls at a level other than EAR99/AT.

Price Protection

Licensee who purchases volume license of the Software allowed to order additional copies with the same price during following period after initial deal: 3 months for 3 to 5 copies, 6 months for 6 to 10 copies and 9 months for 11 or more copies.

Trademarks information

DTM Data Scrubber is trademark of DTM soft.

Licensee hereby grants DTM soft a right to use company name or trade names solely in connection with the rights granted to DTM soft pursuant to marketing materials and web site. If this clause breaches company policy DTM soft is happy to remove it upon request.

What differences between the demo and full versions of DTM Data Scrubber?

General functions

- Demo version allows the user to create not more than 5 rules per project.
- Demo version makes not more than 20 changes per rule during execution.

Supplemental functions

- SQL console partially replaces result values to DEMO string.

No other demo limitations are present except nag-screen at program shutdown.

If you have some question or unusual problem feel free to contact the DTM Data Scrubber technical support at support@sqledit.com

When you contact technical support, you should be prepared to provide the following information:

- DTM Data Scrubber version (you can find this information from About menu item of Help menu).
- Type and version of the ODBC or IDAPI driver or OLE DB provider.
- DBMS version and operating system version (including service pack version, if applicable).
- DTM Data Scrubber Log file.
- A description of what you do before the problem occurs.
- Error messages you see when the problem occurs.
- Your name, company name and how to contact you.

See Also: [log_file](#)

To run installation program:

- Open the windows Start menu and select "Run" item
- Select or enter installation file name and path (scr.exe or scr_d.exe)* and click OK

* - professional and enterprise editions of the tool may have another suffix.

Notes:

- Please be sure that existing version of the tool is not running when you install a new version.
- We recommend to [uninstalling](#) old version of the DTM Data Scrubber before new version installation.
- Installation by administrator for another user is supported for most environments.



Uninstall the Software

The Uninstall feature removes all installed DTM Data Scrubber components and all records in the Windows registry made by the installation script. You can uninstall this program by selecting the "**DTM Data Scrubber**" item in "**Add/Remove Programs Dialog**" in "**Control Panel**".

Another uninstallation way is to run "unins000.exe" from the product's folder directly.

Important! Uninstall feature of the program does not remove files and objects created by users such as configuration files, registry records etc.



How to upgrade your copy of DTM Data Scrubber?

The user can refer to "Check for Update" features to get information about available updates.

Please contact our support staff at support@sqledit.com to upgrade commercial version of the tool.

Demo version is available for [download](#) free of charge.

When you upgrade your copy of DTM Data Scrubber please send us the following information:

- You name, company name and how to contact you
- Payment information (at least "ORDER No" and "Date")



Trademark Information

Microsoft Excel, Microsoft SQL Server and Microsoft Windows are trademarks of Microsoft Corp.

Oracle is a registered trademark of Oracle Corp.

PostgreSQL is copyrighted by PostgreSQL Global Development Group.

Borland Interbase is a registered trademark of Borland Software Corp.

[DTM SQL Editor](http://www.sqledit.com/editor) (www.sqledit.com/editor) is a set of powerful database management tools that allow you to achieve two goals - to have unified access to different types of databases and to have a set of solutions that makes processing your data easy. DTM SQL Editor gives database users, developers and administrators an ability to access different databases, whether desktop or client-server ones (provided you have ODBC driver installed). This is very convenient, since most organizations use several different types of databases installed and each stores data in different formats and with varying parameters. Having a program that can get data from various sources is often essential. Furthermore, in addition to letting you quickly switch between different data sources, DTM SQL Editor lets you see database schema and results of the query execution.

[DTM Migration Kit](http://www.sqledit.com/mk) (www.sqledit.com/mk) is a powerful yet simple data migration tool that comes in handy if you run multiple databases. Use it to import, export or migrate data between different data sources (ODBC, OLE DB, or Oracle Call Interface supported). The program is fully automatic and supports all popular database formats. Simple visual interface lets you set own transformation and flow control rules to give you added flexibility.

[DTM Schema Reporter](http://www.sqledit.com/sr) (www.sqledit.com/sr) is a reporting tool for database schema. The program creates reports in RTF, HTML, XML or plain text formats and supports all common database interfaces - ODBC, OLE DB, or even Oracle Call Interface. This utility helps technical writers and database administrators create a report of any complexity level within seconds. Also, you can alter table order in the report and manually add annotations to the individual tables.

[DTM Data Generator](http://www.sqledit.com/dg) (www.sqledit.com/dg) is a simple, powerful and fully customizable utility that generates data for database testing purposes. Currently, database developers and administrators often have to spend hours of dull work to create test data sets before examining database performance. This tool makes all this unnecessary by automatically creating database objects AND sets of SQL statements, if necessary.

[DTM Data Editor](http://www.sqledit.com/de) (www.sqledit.com/de) is a data viewer and editor for database professionals who are tired of wasting their time on mundane tasks. The program uses form-based interface and works with any ODBC data source. SQL statements are generated automatically and can be modified later. For data that has foreign key - primary key relation, there are options to enter values manually or select them from a list, which is much faster.

[DTM DB Stress](http://www.sqledit.com/stress) (www.sqledit.com/stress) is a utility for stress testing the server parts of information systems and applications, as well as DBMSs and servers themselves. This tool allows you to create and configure a continuous set of requests to the server of the OLAP (query execution) and OLTP (adding, modifying and deleting data in the database) types. At the same time, the user can flexibly change both the number and the priority of this or that type of requests to a database or an application.

[DTM Data Modeler](http://www.sqledit.com/dm) (www.sqledit.com/dm) is a CASE tool for database developers that supports both forward and reverse engineering. It is an easy-to-use tool allowing you to work both with logical and physical data models in the form of an entity-relationship diagram. The product is intended for database architects and developers and works with data sources via the ODBC interface, which means compatibility with all modern DBMS. Along with basic model properties (sets of entities and relationships between them), the program allows you to create indexes and triggers on the physical level corresponding to the tables of the database that is modeled.

[DTM Data Scrubber](http://www.sqledit.com/scr) (www.sqledit.com/scr) is a set of intelligent tools for data verification (audit) and scrubbing (cleaning). Depending on user-defined rules and data properties, the program either creates a report about the actual state of affairs or performs database data correction.

[DTM Data Comparer](http://www.sqledit.com/dcmp) (www.sqledit.com/dcmp) is a visual tool for data compare and synchronization. The program successively views the contents of both tables basing on the order of ascending of unique key values and shows differences or creates synchronization script.

[DTM Schema Comparer](http://www.sqledit.com/scmp) (www.sqledit.com/scmp) is a tool for database schemas comparison and synchronization. The comparison process supports tables, views, indexes, triggers and stored procedures. The visual representation of database schemas as a tree makes the comparison process more comfortable.

[DTM Query Reporter](http://www.sqledit.com/qr) (www.sqledit.com/qr) is a reporting tool for database query. This utility helps technical writers, developers and database administrators create a report based on database query within seconds.

[DTM Schema Inspector](http://www.sqledit.com/si) (www.sqledit.com/si) is a database schema browsing and management tool that let you work with database schemas more effectively.

[DTM DB Event](http://www.sqledit.com/event) (www.sqledit.com/event) is a database monitoring and management tool. This utility allows the user to define a few situations (events). For each event the user can define what the program should do if the event is occur.

[DTM Flat File Generator](#). Easy to use tool that helps any developer or QA engineer to create test data file. It supports tab-delimited, CSV, fixed width and custom separated output files. The generator has powerful import and export file structure features.

[DTM Test XML Generator](#). The tool is powerful generator for XML documents with structure defined by user and random but realistic data. More than 30 predefined generators with powerful pattern engine. The rich import XML structure options are available.

[DTM Data Generator for Excel](#) is a tool for text Excel spreadsheet population. Easy to use interface based on predefined generators, rich value library and high performance.

[DTM Data Generator for JSON](#) produces JSON files with defined structure in a bulk manner. Fast and easy structure editor and smart import options helps the user to generate test set in a few clicks only.

[DTM Database Content Analyzer](#) is a statistical tool for database content. It collects a few dozens of most interesting data: database objects size, value frequency, clusters, etc. This tool replaces and extends "Statistics" report of obsolete versions DTM Schema Reporter.

Database catalog - The collection of system tables, tables that store metadata about that specific database.

Database record - one row in a table (table can be a result of SQL-query).

Database schema - logically connected, usually owner-based, set of DBMS objects (tables, views, procedures etc).

DBMS - database management system.

DBMS connection - the fact that both client and DBMS server have signed a contract and ready to query and data communications.

Drag-n-drop - the file manipulation technique when the mouse is used to move the file from the place of storage to the program, which performs processing.

SQL language - the declarative language used to manipulate the data and its' structure in the modern DBMS and their client applications.

IDAPI - Integrated Database Application Program Interface, unified DBMS access interface.

OCI - Oracle Call Interface, access interface for Oracle Server.

ODBC - Open Database Connectivity, unified DBMS access interface.

Metadata - information about data. See also: database schema

SQL statement - single SQL operator having the complete role in a data manipulation script.

SQL-server - program or program complex, which is able to execute the SQL-queries.

