

StealthTrace Guide

Version 5.1

(C) Eugene Karataev
mailto:support@minimdb.com
<http://www.minimdb.com>

October 27, 2011

Contents

1	StealthTrace components	2
2	StealthTrace installation	3
3	StealthTrace Storage creation	4
4	StealthTrace indices	4
5	Cube definition	5
6	Data loading	7
7	StealthTrace query	8
8	Association Rules	9
9	Appendix 1. StealthTrace type's formats	11
10	Appendix 2. Demo Home Money	12
11	Appendix 3. Demo Market Basket Analysis	13

1 StealthTrace components

StealthTrace is a package to store data and make queries by user with output query results in a cross table.

The package consists of server modules and client-side programs. Server modules implement mechanisms for data storage, a special indexing, querying and issuing of results to client programs.

Client programs consists of two part - StealthTrace for end-user query creation and running and StealthTrace Meta as special tool to define StealthTrace cubes and generate MUMPS code to add data into the cube.

On the server part package use special StealthTrace Storage, which includes special routines (in compiled bytecode) and special data structures.

One MiniM database can contain one StealthTrace Storage. One StealthTrace Storage can contain several cubes and queries.

StealthTrace application is indended to end users to define data selection and how to show reports. StealthTrace Meta application is intended to developers and administrators of data warehouse.

StealthTrace application can export report into portable data file formats and place data into the clipboard to exchange data with other software.

Moreover, StealthTrace can print report with apportion by pages by verial and horizontal. Result of queries in many cases may use several pages by horizontal and printed pages may be pasted together to get full big report.

Server and client parts of StealthTrace contains special algorithms and data structures to use big data volumes on the server and generate big reports.

Server and client parts of StealthTrace may work on one computer and on the different. Client programs connects to the server using special protocol over tcp/ip. Data exchange protocol is supported by StealthTrace and MiniM Database Server. Administrator must provide correct tcp/ip connection only.

StealthTrace Storage - whole routines on the server side and special data in the special formats. StealthTrace programs use only data whithin StealthTrace Storage.

StealthTrace Dimension - conventional axis of the conventional space, where may be defined place of the cube's fact.

StealthTrace Measure - own fact's characteristic, may be used in arithmetic functions to calculate reports.

StealthTrace Fact - one record, unites values by dimensions and measure values.

StealthTrace Cube - union of homogeneous fact records. Cube is conventional name, this name means that the position of the fact is determined by several dimension values. Instead of name Cube many software use special name Hypercube.

StealthTrace Query - definition of the query to the StealthTrace cube including conditions of data selection, what functions must be evaluated and how result must be shown.

2 StealthTrace installation

StealthTrace package requires installed and runned MiniM Database Server version 1.10 or later and installed client side of StealthTrace on each work place.

StealthTrace installer sets up all need client programs, this documentation, compiled routines to create StealthTrace Storage automatically by need and sample StealthTrace demo Home Money.

After installation developer or administrator must run StealthTrace Meta program and create connection to MiniM database, where must be created server-side part of StealthTrace. StealthTrace Meta program automatically check server side configuration and offer to create and initialize StealthTrace Storage by need. This procedure does not requires any other special data or routine import and any procedures execution - all steps are made automatically.

On confirmation StealthTrace Storage creation StealthTrace Meta program automatically imports routines and creates initial data configuration to store cubes definitions, data, indices, queries and all others.

StealthTrace Storage can be created only one in one MiniM database and may be in each MiniM database.

Full program set on the client side is required only for administrators and developers, and for end users recommended to remove stmeta.exe after installation to avoid accidental editing.

3 StealthTrace Storage creation

Before MiniM database and StealthTrace Storage creation administrator must understand some principles of data warehousing and take into account if they are important for task accomplishing.

Most important principles are:

1. Working with big data volumes
2. Data in warehouses are secondary and takes from primary data sources
3. Data in warehouses adds periodically
4. Data warehouses use many read operations

So, the following tips may be recommended for data warehouses:

1. Use separate MiniM database for StealthTrace Storage
2. Increase MiniM global's cache
3. Disable database journaling for warehouse database
4. Use separate hardware server or separate MiniM instance

In any way administrator can use any of available MiniM database, and StealthTrace package will execute need tasks.

One StealthTrace Storage can contain one or more StealthTrace cubes. And cubes may use the same dimensions. Current StealthTrace version have not possibilities to make queries to two or more cubes in one query, but this possibility is planned to the future. So, may be recommended store cubes with the same dimensions in one StealthTrace Storage.

4 StealthTrace indices

StealthTrace Storage uses classic bitmap indices. To determine which records have selected value, for each value are stored special sequences of zero and units. Unit place number is internal identifier of StealthTrace record. And zero means that record with this number have not this value.

To enum all records with selected value StealthTrace Storage get the index sequence and enum all available units. For each found unit StealthTrace Storage calculates unit's position and this number is treated as a record number.

To make logical OR and AND operations StealthTrace Storage make bitwise operations over selected index sequences. Result is a bitmap sequence with records which match logical condition. Bitwise operations are made fast on low level, so search is made much faster than using traditional b-tree indices.

Bitmap indices are much usable to make logical operations over big records set, but traditional indices may be faster to select only several records.

StealthTrace Storage logically make space excision from total StealthTrace cube using the following rules: conditionals by one dimension unites with OR rule and conditions by different dimensions unites by AND rule. So, user can select from total StealthTrace cube several spaces.

All internal index storage, data normalization and index actualization StealthTrace Storage make automatically.

5 Cube definition

StealthTrace Cube contains set of the records or facts. Cube is a combination of Dimensions (conventional axis) and Measures.

Cube axis is a conventional one-dimensional space, in which fact may be positioned and differs from other facts or have the same coordinate.

To understand the conditional axis we will describe a brief overview of mathematics. Let us have some objects. Until we have not defined them above any operations, it just sort of conventional pile.

If each of the object we can enumerate conditioned by certain numbers or name or another in some way, after the intention of such an operation, this pile of objects have become a multitude of objects.

If our conventional numbers or names, which we ascribe to these objects, we can add a rule to organize them and to indicate what number or name must be conditioned to the left or right which (greater or lower), those numbers have already formed space, and our multitude will be an ordered multitude.

If to our provisional numbers or names assigned to objects, we will add rule for determining the action to an arbitrarily chosen pair of numbers or

names, or distance between two objects, or in other words, the addition operation, we already enter the geometry on this space. Traditional operations of addition invariant is the notion of measure zero or null object.

If plus to ordered measure space and the addition operation we will add multiplication of two any selected numbers or names, or actions, then we obtain algebra on this set of numbers or names. Traditional multiplication invariant is the notion of a single object or unity.

Thus, difference of the geometry of space and the algebra are conditional and is determined only by what actions we have defined, and, formally, the algebra is the extended geometry.

To use StealthTrace package it does not need to define geometry and algebra, it is enough to define simple ordered multitude. The values of axis can use only comparison operations and to make StealthTrace queries it is enough to understand what dimension values are greater or lower by selected axis (dimensions).

StealthTrace package use only two operations by sorting - sorting as a strings and as a numbers. Numbers are sorted in arithmetic order and strings are sorted in alphabetic order.

Fact or one record may be at one time in several spaces, for example computer connection to the internet provider may be in the following spaces: City, Street, House number, Date, Time, Connection type, Team number.

The fact's measure is an owned fact's characteristic, and intended to make arithmetic operations, for example it may be conditional quality of connection or nominal speed or number of connected computers.

If report requires to select data by measures (not only by dimensions), it is need to add the same values of measures into specially defined dimension.

Traditionally, data warehouses are includes dimensions of value that can be conventionally enumerated and the number of distinct values are not so great. A value indicators can be arbitrary, including fractional values, and the number of different values of the indices are usually very large, including the case that each of the of facts has its own index value.

StealthTrace cube definition is made by StealthTrace Meta program. Traditional sequence of cube definition may looks like the following:

1. Developer creates the cube and define dimensions and measures.
2. Developer uses automatically generated by StealthTrace Meta code to add one record in own routine to add data to the cube.

3. Developer exports cube definition from StealthTrace Meta and exports set of routines for data transfer and sends full set of files to warehouse administrator.
4. Administrator runs StealthTrace Meta and creates by need initial StealthTrace Storage.
5. Administrator imports cube definition by StealthTrace Meta and imports and compiles set of routines which adds data.
6. Administrator set up need procedures defined by developers for periodicity of data addition.

In many cases for small warehouses developer and administrator can be the same person and developer and production servers may be the same computers.

6 Data loading

Mainly data warehouses are data storages with periodic data addition. In a time storages accumulates more and more data and data have a great value for an analysis.

Data filling of StealthTrace Storages are made record-by-record, and must be passed denormalized values of dimensions and measures. StealthTrace Storage automatically make data transformations and internal normalizations by needs. StealthTrace Storage use internal numbers by needs.

StealthTrace cube developer must run StealthTrace Meta program, open cube and look menu item Meta | Code to Add Data, and get code template for records addition. This code template can be used as code snippet in custom routine with data additions.

Code for data additions is a code for values transfer in special order to StealthTrace Storage. Code transfer values of dimensions and measures in depending of internal cube definition.

StealthTrace Storage on import of cube definition supports the same dimensions and measures order to made developed routine still work on the other server. If cube was created not by import cube definition developed on other server, dimensions and measures order may differs and coherence does not guaranteed.

Developer must save in own routine the same fields order. If developer change cube definition, code for data addition must be recreated by current cube definition and developer must check fields order again.

Moreover, developer must define data source from which data must be transferred and defined the method how distinguish already added data and what portion of data must be added now. It may be any data source - external file in a special format, special application which call MiniM, ODBC data source, other MiniM database or any other data source.

For example, it may be entire data file import, or a call to MUMPS-writted application to call by data after last update time.

In some cases data may be accumulated from different data sources. Information which is actual for analysis, may be absent in OLTP database but may be stored in external databases or dictionaries.

Developer must create routines in a way, that developed routines may be exported and imported in a production system. With cube definition file and set of routines developer must create special instructions (by need) to set up external data gateways or use data gateway conventions described by administrator.

Moreover, developer must use special conventions to normalize values of dimensions and measures. StealthTrace Storage distinguish different values as a different values. For example, city names must be in lower case with upper first letters, without leading and trailing space symbols, without doubled space symbols, etc. In any way, if two values means the same, this values must be identical.

If the data to load is guaranteed obtained from the same source and their values are already normalized, guaranteed, some of the work on normalization of the values the developer can drop and can trust to source of data.

The overall result of the developer is the package of files that may be passed to the server administrator, to import and run. It is recommended to avoid certain magic globals, magic identifiers, or system settings whose values may be different on a different servers or may be absent, due to which data can come in incorrect state in terms of analysis.

7 StealthTrace query

StealthTrace queries are created, edited and runned by StealthTrace program (st.exe).

For new query must be selected one of available StealthTrace cube on the server in the current StealthTrace Storage. And for this cube must be selected axis (dimensions) for left dimension's set and top dimension's set. For example, if user selects City as left dimension and Year and Month as top dimensions, report will show data of City by Year and Month. In dependence of dimension's order report may detailize Year by Month or Month by Year.

If user does not create any conditions for query, report will be made by all available in the cube data records. User can excise some parts of total data by conditions defined for cube dimensions. StealthTrace use OR rule for conditions by the same dimension and AND rule for conditions by different dimensions.

Axis defines how data must be shown, conditions defines which parts of total cube must be shown and selected measures defines what must be shown for selected records. User must select at least one cube measure and functions which must be applied. User can select one or many measures, all selected measures will be included into report.

If user specify option to calculate subtotals for dimension, StealthTrace will calculate subtotals using the same function of function for measure. For example, if user select function Summary, subtotals will contain Summary too, if select Middle, subtotals will contain Middle, and subtotals will contain data by appropriate group of data.

StealthTrace report by query show color highlightning for cells with different mean. This colors user can set up by need in program options.

StealthTrace query is automatically stored in the database and may be used later or may be edited. StealthTrace does not store reports on the server and all reports are made each time by current cube data.

Reports may be exported to widely used data formats to transfer data to other applications or reports may be printed. Moreover, reports may be exported to HTML format to view by browser later without connection to database.

StealthTrace reports after creation allows select number of decimal digits to view report in a wanted form.

8 Association Rules

Application StealthTrace Association Rules executes data analysis, searches and shows association rules.

First, need to select one of available StealthTrace cube and define rules parameters - premises, consequences, select and join conditions, rules support and confidence percents.

If was not specified selection conditions, application use all data in the cube, otherwise analyse only data meets to conditions. Selection conditions are conditions to values by cube axis. On selection application join several conditions by one axis with OR and by different axis with AND rule.

Application searches rules by specified premises. Need to be specified at least one axis as a premise and at least one axis as a consequence. If was specified several axis as a premises and consequences,, application searches this combinations.

Fact joins or list of axis to join are used to search correlations between different facts, joined by specified axis values. If joins was not specified, application can analyze only correlations between different dimensions. And, if joins are specified, can analyze the same dimensions for facts with the same joins values.

To typical association rules analisis commonly refer two tasks: search associations between different dimensions and associations between the same dimensions with joins by other dimension.

For example, first task is searching association what goods are preferably sells by weekdays or by day time. In this case premises are week day or day time and consequence are goods.

And, for example, second task is a classic Market Basket Analysis task with searching what goods associates with other goods. In this case premise and consequence are good's dimension and joins are check numbers. This join need to get association by one check. To make this analisis, StealthTrace cube must be created to contain need dimensions and information. Analisis can be made with joins not only by checks, also can be used shop number, day, or day time, or any other what want to see analyst.

StealthTrace Association Rules comes with demo example MarketBasket to demonstrate how to analyse cube and search association rules.

Much important to understand analisis are values of rules support and rules confidence. Rule support is a percent of found premise within all premises. If this percent is big, this premises are widely encounters. And vice versa, small percent mean that this premise are infrequent. Rule confidence is a percent of combination of pairs of premise and consequence within all this premises. If this percent is big, this rule is much confidence and

hardly confirms, and vice versa, if this percent is small, this rule confirms infrequently.

To understand rules analysis, the support value characterises how widely this rule encounters in data, or how this rule is obvious. The confidence value characterises how this rule is stable. So, application have filter for both percents to limit top and bottom values to hide rules with very big and very small values of support and confidence. This allow not to show obvious and uninteresting rules.

Association rules options are automatically stored in the database and may be used or changed later. Rule reports are not stored on the server and all reports are created on the fly only by current StealthTrace cube content.

Reports may be exported to widely used data formats to transfer data to other applications or reports may be printed. Moreover, reports may be exported to HTML format to view by browser later or to PDF without connection to database.

To search association rules is used separate application star.exe with special user interface for association rules.

9 Appendix 1. StealthTrace type's formats

StealthTrace records must have the following formats and follow encoding conventions:

1. String - string as is.
2. Number - number in the MUMPS notation, with possible decimals digits.
3. Date - value of the date in the \$HOROLOG system variable format.
4. WeekDay - integer number from 1 (Monday) to 7 (Sunday).
5. Week - integer number from 1 (first week of the year).
6. Month - integer number from 1 (January) to 12 (December).
7. Quarter - integer number from 1 (first quarter) to 4 (fourth quarter).
8. Year - integer number of the year. Recommended to use 4-digits format.
9. Currency - number in the MUMPS notation, with two possible decimal digits.

10 Appendix 2. Demo Home Money

StealthTrace package contains a demo Home Money. An example describes the household finances of a family. An example is in a subdirectory /demo and consists of two files with the definition of cube (HomeMoney.stm) and with routine for generating random data to it (HOMEMONEY.rou).

Example describes the expenses of family members, for every fact information explains by dimensions - Account, Person, Year, Month, Day, Type and by measures Bill amount, Discount and Payment.

To run the sample, perform the following steps:

1. Select the database where need to be created StealthTrace Storage and cube.
2. Run StealthTrace Meta (stmata.exe). On start this application check that the database contains StealthTrace Storage. If not, application offer to create and initialize StealthTrace Storage.
3. From the menu item Meta | Import Metadata import file HomeMoney.stm from the /demo subdirectory.
4. Close application StealthTrace Meta.

After this steps database will contains StealthTrace Storage and Home Money cube definition. To fill this cube by data records perform the following steps:

1. Run MiniM Routine Editor or MiniM Control Center and connect with the same database. Import routine from the file HOMEMONEY.rou from the /demo subdirectory with compilation.
2. Run MiniM console or telnet, switch database and execute the following command:

```
d CreateRecords^HOMEMONEY()
```

After that, the cube Home Money will contains 2,000 records with randomly generated values.

And, now may be run end-user StealthTrace application to create and execute cube queries, for example:

1. Run end-user application StealthTrace (st.exe)
2. Create new query (for example, Home Query), and select cube Home Money.
3. Next may be selected query options, for example Left Dimensions: Person; Top Dimensions: Year, Month; Query Condition: Account = Transport; Measures: Payment (Summary).
4. Run query.

By changing the query parameters, you can explore the data cube in different Home Money views.

11 Appendix 3. Demo Market Basket Analysis

StealthTrace package contains a demo Market Basket Analysis. An example describes sells in some shop. Demo example is in the /demo subdirectory and consists of two files MarketBasket.stm with StealthTrace cube definition and MBASKETAN.rou with routine to add data about sells.

This demo contains only short variant of the sells, only about data to search rules. Cube includes dimensions Selling date, Selling week day, Selling check number and Selling good. Cube contains only one fake measure with count of cells and all facts have the same value 1.

It is important note, that the cube for rules analisis can differ from cube for other analisis tasks. In particulars, for classical Market Basket Analysis the check number is required. In the case of two shops can have the same check's numbers, need to be used joins by shop number and check number. Also, frequently, analisis makes not by good name, but by some good'd class. For example, all milk assortment can be united to Milk. Or, to analyze correlation of day time and milk volume, need to have milk package class as a dimension.

To run the sample, perform the following steps:

1. Select the database where need to be created StealthTrace Storage and cube.

2. Run StealthTrace Meta (stmata.exe). On start this application check that the database contains StealthTrace Storage. If not, application offer to create and initialize StealthTrace Storage.
3. From the menu item Meta | Import Metadata import file MarketBasket.stm from the /demo subdirectory.
4. Close application StealthTrace Meta.

After this steps database will contains StealthTrace Storage and Market Basket cube definition. To fill this cube by data records perform the following steps:

1. Run MiniM Routine Editor or MiniM Control Center and connect with the same database. Import routine from the file MBASKETAN.rou from the /demo subdirectory with compilation.
2. Run MiniM console or telnet, switch database and execute the following command:

```
d ^MBASKETAN
```

After that, the Market Basket cube will contain several records with demo data about sells.

Next can be run StealthTrace Association Rules application star.exe to define and search association rules.

As an exmple how to search rules of goods association select as a premise and consequence the Selling Good dimension and Selling Check as a join.

Application shows all found rules matched to specified conditions. Application show report about found rules in two form - as full list of rules and as a What-If report. On switch report to What-If mode application shows all found premises and for each premise show all consequences for this premise. Colimns may be sorted by ascending or descending.

In particulars, demo example show that the bacon associates with the potatoes, cookies with the milk and beer with the chip.