

Understanding the SAS/MDDB® Server to Process Large Files



Prepared by Dana Rafiee

Destiny Corporation • 1321 Silas Deane Highway • Wethersfield, CT 06109-4302

Phone: (860) 721-1684 • 1-800-7TRAINING • Fax: (860) 721-9784 • Email: destinycorp@compuserve.com

Web Site: www.destinycorp.com

Copyright © 1998 Destiny Corporation. All Rights Reserved.

What is an MDDB?

A new method of containing summarized information in a read only format has been introduced in the SAS System. This is the SAS System MDDB.

Each MDDB contain certain characteristics:

- at least one NWAY summary table.
- a file type of MDDB
- unlimited subtables
- unlimited classification variables
- unlimited analysis variables
- a maximum of 8 stored statistics
- up to 13 other statistics (available as a combination of the 8 stored statistics)
- no missing cross tabulation data
- an incremental update capability
- online and batch creation capability

The NWAY table is similar to the NWAY output produced for the PROC summary step.

The subtables produced are equally similar to specific `_TYPE_` records produced as a consequence of the PROC summary step.

Available Statistics

The 8 stored statistics are:

Statistic	Meaning
N	Number of Observations that have non -missing values.
SUM	Total of all the values within a given variable.
SUMWGT	Sum of the WEIGHT variable values.
UWSUM	
NMISS	Number of Observations that have missing values.
USS	Value of the Uncorrected Sum of Squares.
MIN	The minimum value of a given variable.
MAX	the maximum value of a given variable.

The 13 available statistics (as a combination of the 8 stored statistics) are:

Statistic	Meaning
AVG	The average of all the values of a given variable.
RANGE	the range between the min and max values of a given variable.
PCTN	Frequency Percentages.
PCTSUM	Summary Percentages.
CSS	Value of the Corrected Sum of Squares.
VAR	Value of the Variance of a given variable.
STD	Value of the Standard Deviation of a given variable.
STDERR	Value of the Standard Error of the mean for a given variable.
CV	Value of the coefficient of variation expressed as a percentage.

T	Value of the Student's t value for testing the hypothesis that the population mean is zero.
PRT	The probability of a greater absolute value of the Student's t value under the hypothesis that the mean is zero.
LCLM	Value of the Lower Confidence Limit
UCLM	Value of the Upper Confidence Limit

The Terminology and Structure of an MDDB

An MDDB is a collection of subtables.

A subtable is a collection of summarized data.

The original data that was used in the creation of the MDDB is known as the BASE TABLE. The base table may be single file or a view containing many tables and computed tables.

It is possible to drilldown through the data. The DRILLPATH is the route that the drilldown will take. For example, a DRILLPATH could be state, city, road, and hotel.

A variable that is within the given data can be defined as a CATEGORICAL FIELD if it defines groups within the data such as Region or Grade.

If an MDDB contains a high percentage of intersections with missing data then the multi-dimensional data set can be defined as SPARSE.

A LEVEL is the number of unique values within a given variable. For example, there would be 50 levels to a variable containing the State codes for the United States.

The data will contain the following fields. CLASSIFICATION which are levels. ANALYSIS which are measures. NWAY, which are subtables, made up from classification and analysis variables.

Before a user attempts to create an MDDB they should a good understanding of the following points:

- The problem they hope to solve.
- The data being used and how it will be used.
- The expectations of the clients awaiting the end MDDB.

The MDDB Procedure is made up of possible 4 statements.

They are:

Proc Statement

- with 4 further possible options.

Out =, the output data set. The only required option.
 Data =, the input data set. If none is specified then `_LAST_` is used.
 In =, used for incremental updates.
 Label =, the description of the MDDB. Maximum of 80 characters in length.

An example of the Proc Statement could be:

```
proc mddb data=mddbserve.demog out=mddbserve.mddb
label='Example Mddb';
```

Class Statement

- a list of the variables to be used in the Mddb. The listed variables can be numeric or character. There is the option of setting the order of the variables; the default order is ascending.

Order options are:

ASCENDING	sorted in ascending order.
DESCENDING	sorted in descending order.
ASCFORMATTED	sorted in ascending order by formatted value.
DESFORMATTED	sorted in descending order by formatted value.
DSORDER	sorted in the order that they occur in the data set.

The sort order must be specified after a '/' which follows the variable name.

To specify a different sort order for different variables use multiple class statements.

An example of the Class Statement could be:

```
class gender/descending;
class age;
class height/ascformatted;
class grade/dsorder;
class weight/desformatted;
```

NOTE: age would be sorted in ascending order.

Hierarchy Statement

The hierarchy statement creates named subtables within the Mddb. Within the SAS System named subtables are by default assumed to be non-display hierarchies. If the do not name your hierarchy the SAS System will give it a name of HIERn where n is from 1 to the number of requested hierarchies.

To specify a name with the hierarchy use the name= option. The name must be in quotes if it contains any spaces.

To overwrite the non-display default provided by the SAS System use the display= option. The available values for this option are YES and NO.

The order of the separate hierarchy statements is not important but the order of the variables within the same statement is important if the display type is YES.

An example of the Hierarchy Statement could be:

```
hierarchy gender weight height/name='Example Hierarchy'
display=YES;
```

VAR Statement

The VAR statement lists the variables that are to be used as analysis variables. All the listed variables must be numeric. The statistics to be used with the variable must be specified after the variable and must be one of the 8 stored statistics. We have already described the available statistics.

An example of the VAR Statement could be:

```
var height / n sum avg;
var weight / min max range;
```

We can now put all the statements together to complete the Mddb procedure.

The Complete Mddb Procedure

```
proc mddb data=mddbserve.demog out=mddbserve.mddb
label='Example Mddb';
class gender/descending;
class age;
class height/ascformatted;
class grade/dsorder;
class weight/desformatted;
hierarchy gender weight height/name='Example Hierarchy'
display=YES;
var height / n sum;
var weight / min max;
run;
```

If we run this program in the SAS System we will create the file mddbserve.mddb. We will now see how to view the file.

Locating an Mddb

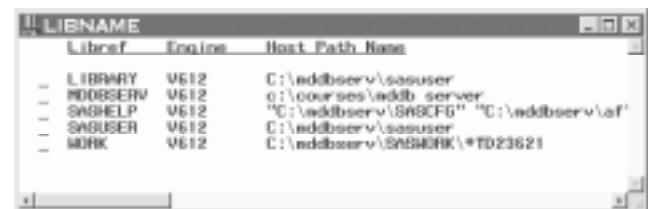
As stated before an Mddb within the SAS System will have a file type of Mddb.

It is possible to check for the existence of an Mddb by using the LIB and VAR windows.

Type lib on the command line.



A window similar to the one below will appear.



Select the MddbSERV libname by typing an "S" in the available space "_".

A window similar to the one below will appear. Scroll down until you find an Mddb file.



An MDDB file

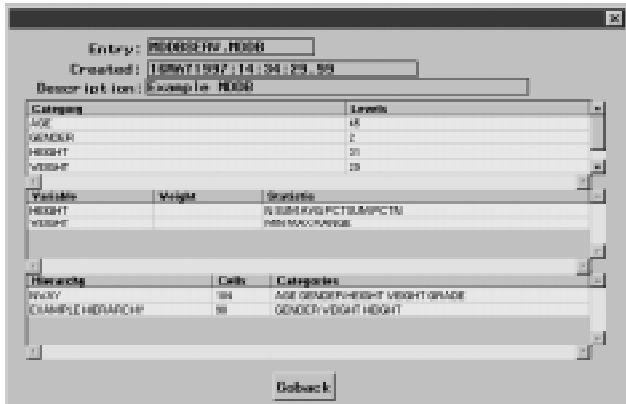
Looking at an MDDB file

Within the SAS System there is a utility window which is available to look at the MDDB file. By using the utility window it is possible to see the:

- general information about the file, such as creation date.
- classification variables and the number of values.
- analysis variables and associated statistics.
- hierarchies (subtables), number of cells and the class variables in the hierarchy.

To view the MDDB file type an “S” in the available space “_”.

A window similar to the following will appear:



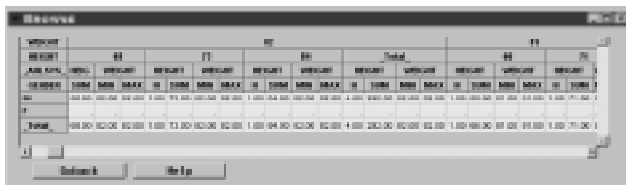
The classification variables showing above are AGE, GENDER, HEIGHT, WEIGHT.

The analysis variables showing above are HEIGHT, WEIGHT. They both have all the associated statistics.

The hierarchies showing above are NWAY, EXAMPLE HIERARCHY. The class variables in the hierarchy are also shown. For the EXAMPLE HIERARCHY the class variables are GENDER, HEIGHT and WEIGHT.

It is also possible to view the MDDB file by typing an “L” instead of an “S” in the available space “_” in the DIR window.

That will give us a window similar to the one below:



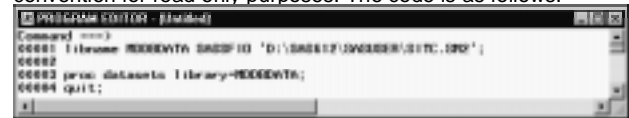
We can see the information contained in the MDDB.

For example the information associated with the * shows that there were 4 males with a weight of 82. The sum of the heights was 292.00.

The right mouse button can be clicked within the table area. The popup list that will appear will contain a list of all the available hierarchies.

Getting Information Out of an MDDB

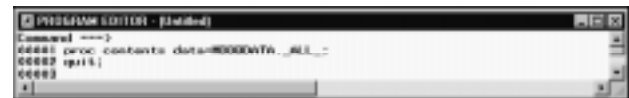
From a Base SAS perspective, quite often there is a need to pull information out of an MDDB. This can be done in base SAS software on most platforms which support the SASSFIO engine. This is an engine designed to allow libname specification to specific locations, including an MDDB. The hierarchies in the MDDB are considered separate data files. These can then be accessed using the standard LIBREF.DATASET naming convention for read only purposes. The code is as follows:



This libname would attach to a specific MDDB. PROC DATASETS would show the hierarchies available as read only SAS data files.



To document the MDDB, a PROC CONTENTS similar to the following could be used.



To achieve the following output.

```

CONTENTS PROCEDURE
-----Directory-----
Libref:      MDDBDATA
Engine:      SASSFIO
Physical Name: D:\SAS612\SASUSER\SITC.SM2
# Name      Mentry Indexes
1  NWAY     DATA
2  HIER1    DATA
3  SITC     DATA
4  CTRY     DATA
5  MONTH    DATA
6  YEAR     DATA
    
```

```

CONTENTS PROCEDURE

Data Set Name: MDDBDATA.NWAY          Observations: 4649
Member Type:  DATA                  Variables: 12
Engine:       SASSFIO                 Indexes: 0
Created:      11:32 Wednesday, April 23, 1997 Observation Length: 77
Last Modified: 11:32 Wednesday, April 23, 1997 Deleted Observations: 0
Protection:                                       Compressed: NO
Data Set Type:                               Sorted: NO
Label:
    
```

-----Engine/Host Dependent Information-----

```

Engine Type:      SAS Cross Platform I/O Engine (READONLY)
Host Platform Family: PC
SAS File Type:    SAS MDDB
File Retrieval Method: Local/Remote Disk
    
```

-----Alphabetic List of Variables and Attributes-----

#	Variable	Type	Len	Pos	Label
10	CTRY	Char	4	69	
7	IETOTVAL	Num	8	48	MIN: ETOTVAL
3	IIVALUE	Num	8	16	MIN: IVALUE
11	MONTH	Char	2	73	
5	NETOTVAL	Num	8	32	N: ETOTVAL
1	NIVALUE	Num	8	0	N: IVALUE
6	SETOTVAL	Num	8	40	SUM: ETOTVAL
9	SITC	Char	5	64	
2	SIVALUE	Num	8	8	SUM: IVALUE
8	XETOTVAL	Num	8	56	MAX: ETOTVAL
4	XIVALUE	Num	8	24	MAX: IVALUE
12	YEAR	Char	2	75	

CONTENTS PROCEDURE

```

Data Set Name: MDDBDATA.HIER1        Observations: 4649
Member Type:  DATA                  Variables: 12
Engine:       SASSFIO                 Indexes: 0
Created:      11:32 Wednesday, April 23, 1997 Observation Length: 77
Last Modified: 11:32 Wednesday, April 23, 1997 Deleted Observations: 0
Protection:                                       Compressed: NO
Data Set Type:                               Sorted: NO
Label:
    
```

-----Engine/Host Dependent Information-----

```

Engine Type:      SAS Cross Platform I/O Engine (READONLY)
Host Platform Family: PC
SAS File Type:    SAS MDDB
File Retrieval Method: Local/Remote Disk
    
```

-----Alphabetic List of Variables and Attributes-----

#	Variable	Type	Len	Pos	Label
10	CTRY	Char	4	69	
7	IETOTVAL	Num	8	48	MIN: ETOTVAL
3	IIVALUE	Num	8	16	MIN: IVALUE
12	MONTH	Char	2	75	
5	NETOTVAL	Num	8	32	N: ETOTVAL
1	NIVALUE	Num	8	0	N: IVALUE
6	SETOTVAL	Num	8	40	SUM: ETOTVAL
9	SITC	Char	5	64	
2	SIVALUE	Num	8	8	SUM: IVALUE
8	XETOTVAL	Num	8	56	MAX: ETOTVAL
4	XIVALUE	Num	8	24	MAX: IVALUE
11	YEAR	Char	2	73	

CONTENTS PROCEDURE

```

Data Set Name: MDDBDATA.SITC         Observations: 2081
Member Type:  DATA                  Variables: 1
Engine:       SASSFIO                 Indexes: 0
Created:      11:32 Wednesday, April 23, 1997 Observation Length: 5
Last Modified: 11:32 Wednesday, April 23, 1997 Deleted Observations: 0
Protection:                                       Compressed: NO
Data Set Type:                               Sorted: NO
Label:
    
```

-----Engine/Host Dependent Information-----

```

Engine Type:      SAS Cross Platform I/O Engine (READONLY)
    
```

```

Host Platform Family: PC
SAS File Type:      SAS MDDB
File Retrieval Method: Local/Remote Disk

-----Alphabetic List of Variables and Attributes-----

# Variable Type Len Pos
-----
1 SITC Char 5 0
CONTENTS PROCEDURE

Data Set Name: MDDBDATA.CTRY          Observations: 179
Member Type:  DATA                  Variables: 1
Engine:       SASSFIO                 Indexes: 0
Created:      11:32 Wednesday, April 23, 1997 Observation Length: 4
Last Modified: 11:32 Wednesday, April 23, 1997 Deleted Observations: 0
Protection:                                       Compressed: NO
Data Set Type:                               Sorted: NO
Label:
    
```

-----Engine/Host Dependent Information-----

```

Engine Type:      SAS Cross Platform I/O Engine (READONLY)
Host Platform Family: PC
SAS File Type:    SAS MDDB
File Retrieval Method: Local/Remote Disk
    
```

-----Alphabetic List of Variables and Attributes-----

#	Variable	Type	Len	Pos
1	CTRY	Char	4	0

CONTENTS PROCEDURE

```

Data Set Name: MDDBDATA.MONTH        Observations: 9
Member Type:  DATA                  Variables: 1
Engine:       SASSFIO                 Indexes: 0
Created:      11:32 Wednesday, April 23, 1997 Observation Length: 2
Last Modified: 11:32 Wednesday, April 23, 1997 Deleted Observations: 0
Protection:                                       Compressed: NO
Data Set Type:                               Sorted: NO
Label:
    
```

-----Engine/Host Dependent Information-----

```

Engine Type:      SAS Cross Platform I/O Engine (READONLY)
Host Platform Family: PC
SAS File Type:    SAS MDDB
File Retrieval Method: Local/Remote Disk
    
```

-----Alphabetic List of Variables and Attributes-----

#	Variable	Type	Len	Pos
1	MONTH	Char	2	0

CONTENTS PROCEDURE

```

Data Set Name: MDDBDATA.YEAR         Observations: 1
Member Type:  DATA                  Variables: 1
Engine:       SASSFIO                 Indexes: 0
Created:      11:32 Wednesday, April 23, 1997 Observation Length: 2
Last Modified: 11:32 Wednesday, April 23, 1997 Deleted Observations: 0
Protection:                                       Compressed: NO
Data Set Type:                               Sorted: NO
Label:
    
```

-----Engine/Host Dependent Information-----

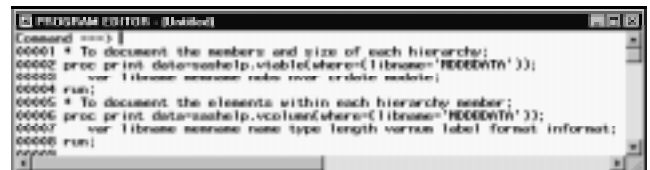
```

Engine Type:      SAS Cross Platform I/O Engine (READONLY)
Host Platform Family: PC
SAS File Type:    SAS MDDB
File Retrieval Method: Local/Remote Disk
    
```

-----Alphabetic List of Variables and Attributes-----

#	Variable	Type	Len	Pos
1	YEAR	Char	2	0

Since data in MDDBs can change and the structure is supported by most aspects of the SAS System, consider using views to help document the metabase structure of what is available, and get direct access to the information.



Yields the following output.

OBS	LIBNAME	MEMNAME	NOBS	NVAR	CRDATE	MODATE
1	MDDBDATA	NWAY	4649	12	23APR97:11:32:45	23APR97:11:32:48
2	MDDBDATA	HIER1	4649	12	23APR97:11:32:45	23APR97:11:32:48
3	MDDBDATA	SITC	2081	1	23APR97:11:32:45	23APR97:11:32:48
4	MDDBDATA	CTRY	179	1	23APR97:11:32:45	23APR97:11:32:48
5	MDDBDATA	MONTH	9	1	23APR97:11:32:45	23APR97:11:32:48
6	MDDBDATA	YEAR	1	1	23APR97:11:32:45	23APR97:11:32:48
I						
L M N						
I E L V F F						
B M E A L O O						
N N N T N R A R R						
O	A	A	A	Y	G	N B M M
B	M	M	M	P	T	U E A A
S	E	E	E	E	H	M L T T
1	MDDBDATA	NWAY	NIVALUE	num	8	1 N: IVALUE
2	MDDBDATA	NWAY	SIVALUE	num	8	2 SUM: IVALUE
3	MDDBDATA	NWAY	IIVALUE	num	8	3 MIN: IVALUE
4	MDDBDATA	NWAY	XIVALUE	num	8	4 MAX: IVALUE
5	MDDBDATA	NWAY	NETOTVAL	num	8	5 N: ETOTVAL
6	MDDBDATA	NWAY	SETOTVAL	num	8	6 SUM: ETOTVAL
7	MDDBDATA	NWAY	IETOTVAL	num	8	7 MIN: ETOTVAL
8	MDDBDATA	NWAY	XETOTVAL	num	8	8 MAX: ETOTVAL
9	MDDBDATA	NWAY	SITC	char	5	9
10	MDDBDATA	NWAY	CTRY	char	4	10
11	MDDBDATA	NWAY	MONTH	char	2	11
12	MDDBDATA	NWAY	YEAR	char	2	12
13	MDDBDATA	HIER1	NIVALUE	num	8	1 N: IVALUE
14	MDDBDATA	HIER1	SIVALUE	num	8	2 SUM: IVALUE
15	MDDBDATA	HIER1	IIVALUE	num	8	3 MIN: IVALUE
16	MDDBDATA	HIER1	XIVALUE	num	8	4 MAX: IVALUE
17	MDDBDATA	HIER1	NETOTVAL	num	8	5 N: ETOTVAL
18	MDDBDATA	HIER1	SETOTVAL	num	8	6 SUM: ETOTVAL
19	MDDBDATA	HIER1	IETOTVAL	num	8	7 MIN: ETOTVAL
20	MDDBDATA	HIER1	XETOTVAL	num	8	8 MAX: ETOTVAL
21	MDDBDATA	HIER1	SITC	char	5	9
22	MDDBDATA	HIER1	CTRY	char	4	10
23	MDDBDATA	HIER1	YEAR	char	2	11
24	MDDBDATA	HIER1	MONTH	char	2	12
25	MDDBDATA	SITC	SITC	char	5	1
26	MDDBDATA	CTRY	CTRY	char	4	1
27	MDDBDATA	MONTH	MONTH	char	2	1
28	MDDBDATA	YEAR	YEAR	char	2	1

Remember, it is possible to create code that accesses the particular hierarchies similar to the way SAS data sets are read. Consider the following data step code.

```
data work.one;
  set MDDBDATA.HIER1(keep=year month sitc
sivalue where=(year='98'));
  ...
run;
```

SAS, SAS/ACCESS, SAS/AF, SAS/ASSIST, Base SAS, SAS/CALC, SAS/CONNECT, SAS/EIS, SAS/ETS, FRAME, SAS/FSP, SAS/GIS, SAS/GRAPH, SAS/IML, SAS/INSIGHT, SAS/LAB, SAS/MDDB Server, SAS/CPE, SAS/OR, SAS/PH-CLINICAL, SAS/QC, SAS/SHARE, SAS/STAT, SAS/TOOLKIT are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are registered trademarks or trademarks of their respective companies.