

SAS / DB2 Date Compatibility

Miriam W. Rosenblatt

Division of General Internal Medicine and Health Care Research, University
Hospitals of Cleveland

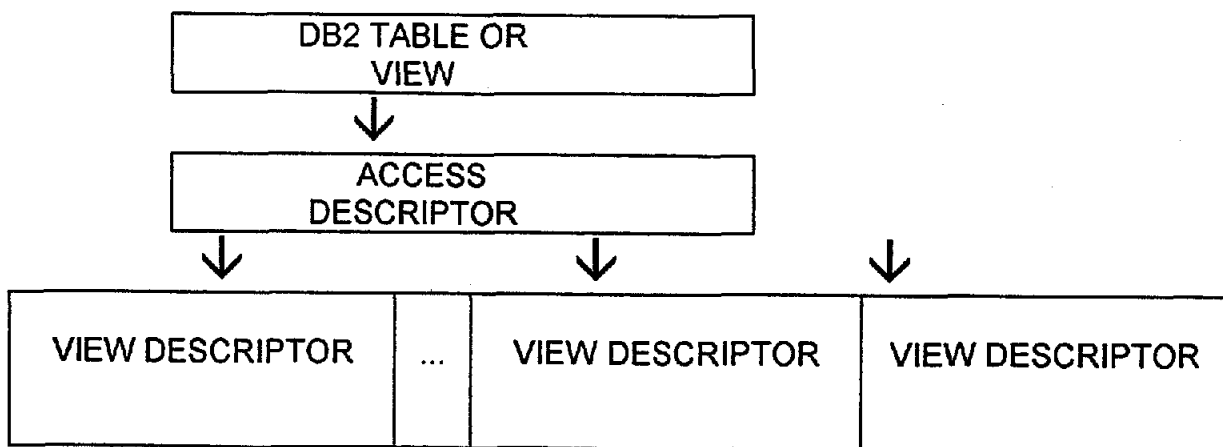
Abstract

This paper describes compatibility differences between date information in a SAS dataset and DB2, and provides basic code to standardize these date values. The SAS/ ACCESS® interface to DB2™ documentation does not adequately address date conversion between the two platforms with regard to access and view descriptors. Problems can occur when you match date values in a SAS dataset and a DB2 file.

This paper is useful for those who combine SAS datasets with data from SAS and any other relational tables using the SAS / ACCESS interface.

Introduction

The SAS / ACCESS interface to DB2 is a powerful interface allowing transparent access to the DB2 relational tables. A primary connection, called an access descriptor, is set up. View descriptors are then derived from the access descriptor. These descriptors are illustrated below.



Descriptors are set up interactively using PROC ACCESS. These view descriptors can then be manipulated and analyzed as SAS datasets with the usual SAS statements and procedures. The procedures are easy to learn and are very straightforward. However, when DB2-derived data are match-merged or compared with variables from a SAS dataset, the variables being compared must be standardized. The SAS documentation does not adequately address the way the date values are defined between the two platforms if the matching variable is a date.

The Problem:

DB2 stores dates as an YYMMDD numeric with a length of 6. SAS stores dates in its own format, represented by the number of days between January 1, 1960 and the given date of interest. When you define both the access descriptor and the view descriptors, you have the option of changing the formats of the DB2 variables. If you change any of the date definitions to a SAS date format, such as YYMMDD6., you will get incorrect results.

The first example below describes a report on the length of stay (LOS) for a group of subjects. The original data is a SAS dataset containing the study ID, hospital number, and admission date. The DB2 file contains the hospital ID, admission date, and the discharge date. The access and view descriptors for the DB2 tables are set up below in Table 1. The data are subsetted based on the start of the study, and limiting the age of the subjects to 70 and older.

Table 1. Access / View Descriptor

Descriptor:	Library:	Member:	Type: Access
Authorization ID: MYID	Table Name: MYDB2TBL	SSID: DB2	
Func	Sel	Column Name	SAS Name Format
—	*	PAT_NUMB	PATNUMB 11.0
—	*	DSCHDATE	DSCHDATE 11.0
—	*	ADM_DATE	ADMDATE 11.0

```
+ Selection Criteria Entry Window -----+
| Command ===>                               |
|                                         |
| Where adm_date >= 930818 and Pt_age >= 70   |
+-----+
```

The program above combines the DB2 data and SAS data in a batch file that generates two kinds of reports: length of stay for study subjects enrolled; report of subsequent admissions.

The first example shows what happens when you change the date values in the access descriptor to a "MMDDYY6." format (Table 2). It may appear simple to merely convert the date variables into a SAS date format. This would allow you to directly compute the LOS by using the date values. For the first type of report, however, there are no results. When the second report is run, both the date values and the LOS are incorrect. Since the SAS dates are different from the DB2 dates in these particular examples, the dates between the SAS dataset and DB2 do not match. If a large file is being used, incorrect dates may merge with the wrong corresponding dates resulting in an incorrect report. The batch program and results are located in **Appendix 1, Output 1**.

Table 2. Incorrect DB2 setup using MMDDYY6.

View Descriptor

Descriptor:	Library:	Member:	Type: VIEW
Output SAS Data Set:	Library:	Member	
Func	Sel	Column Name	SAS Name Format
—	*	PAT_NUMB	PATNUMB 11.0
—	*	DSCHDATE	DSCHDATE YYMMDD6.
—	*	ADM_DATE	ADMDATE YYMMDD6.

+ Selection Criteria Entry Window -----+

Command ==>
Where adm_date >= 930818 and Pt_age >= 70

+-----+

The Solution

The simplest approach is not to change the DB2 values. SAS has many functions that allow easy conversion and data manipulation. Therefore, the date variables should be modified in the SAS dataset and matched with the DB2 dates when the files are joined. Thus, the access and views descriptors should remain exactly the same as in Table 1.

Functions should be used to convert the SAS dates in the batch program. The statement below converts a SAS date into a numeric variable with a length of 6. The new variable now matches the DB2 date variable and can be used to merge both files.

```
NEWDATE = INPUT (PUT (SASDATE, YYMMDD6.), 6.);
```

The batch report modifies the date variables in the SAS file by converting the date values into numeric variables, then joining the DB2 date variables with the modified dates. To compute the length of stay, *the numeric dates must be converted back to SAS dates*. Make sure the length statements are compatible with the variables. There are 0 values in missing discharge data in this particular example. For those cases, you may want to recode those discharge values to missing.

```
SASDATE = INPUT (PUT (NUMDATE, 6.), YYMMDD6.);
```

The correct report is shown in **Appendix 1, Output 2**.

Discussion:

Interfaces between SAS and relational data are supposed to be both user friendly and transparent. While the descriptors are very easy to set up and work with, it is important to understand how the variables in the source data are defined in the relational database system being accessed. The best way to minimize problems when combining both types of data is to leave the definitions alone in the relational data, and use SAS to redefine the needed variables.

References:

1. SAS Institute, Inc. (1989), SAS ® Guide to the SQL Procedure, Usage and Reference, First Edition, Version 6, Cary, NC: SAS Institute Inc.
2. SAS Institute, Inc. (1989), SAS/ACCESS Interface to DB2™, Usage and Reference, First Edition, Version 6, Cary, NC: SAS Institute Inc.
3. SAS Institute, Inc. (1990), SAS Language Reference, Version 6, First Edition, Cary, NC: SAS Institute Inc.

SAS, SAS/ ACCESS are registered trademarks of SAS Institute, Inc. in the USA and other countries. ® indicates USA registration.

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

Acknowledgments

I would like to thank Barbara Juknialis for her editorial advice and Linda M. Quinn for her review of the manuscript.

Appendix 1. SQL Code and Output

SAS Program:

SAS data set ;

```
DATA IDS ;  
  SET DH508.DEMOGWOR (KEEP=SUBJECT HOSPID ADMDATE )  
  ;  
  *** CONVERT DATE TO NUMERIC LENGTH 6;  
  ADMDATE2 = INPUT(PUT(ADMDATE, YYMMDD6.), 6.);
```

SQL: Study subjects only for the first report

```
PROC SQL;  
  CREATE TABLE DB2CHK1 AS  
  SELECT A.PATNUMB, A.DSCHDATE,  
         B.SUBJECT, B.HOSPID, B.ADMDATE  
  FROM MYLIB.LOSEX1 AS A  
       IDS          AS B  
  WHERE A.PATNUMB = B.HOSPID AND  
        B.SUBJECT GE 700 AND  
        A.ADMDATE = B.ADMDATE  
  ORDER BY SUBJECT;
```

```
DATA LOS; SET DB2CHK1;  
  *** CONVERT NUMERIC DATES TO SAS VALUES;  
  LENGTH SASDCDT SASADDT 6;  
  SASDCDT=INPUT(PUT(ADMDATE2, 6.), YYMMDD6.);  
  SASADDT=INPUT(PUT(DSCHDATE, 6.), YYMMDD6.);
```

Readmit SQL: for the second report

```
PROC SQL;  
  CREATE TABLE DB2CHK2 AS  
  SELECT A.PATNUMB, A.DSCHDATE, A.ADMDATE AS READMIT,  
         B.SUBJECT, B.HOSPID, B.ADMDATE  
  FROM MYLIB.LOSEX1 AS A  
       IDS          AS B  
  WHERE A.PATNUMB = B.HOSPID AND  
        B.SUBJECT GE 700 AND  
        A.ADMDATE GT ADMDATE  
  ORDER BY SUBJECT;
```

```
DATA LOS; SET DB2CHK2;  
  *** DISCHARGE DATES OF 0 SHOULD BE MISSING ;  
  IF DSCHDATE= 0 THEN DSCHDATE= .;  
  *** CONVERT NUMERIC DATES TO SAS VALUES;  
  LENGTH READMIT2 DCDATE 6;  
  READMIT2= INPUT(PUT(READMIT, 6.), YYMMDD6.);  
  DCDATE=INPUT(PUT(DSCHDATE, 6.), YYMMDD6.);  
  LOS=(DCDATE - READMIT2) +1 ;  
  FORMAT READMIT2 YYMMDD6.;
```

Output 1: Incorrect Results

1. Report 1: Study Subjects [NO RESULTS]

2. Report 2: Readmissions

SUBJECT	READMIT	DSCHDATE	LOS
700	090731	090803	4
700	331217	331222	6
700	361210	361220	11
700	360918	360922	5
700	611114	611117	4

N=5

SUBJECT	READMIT	DSCHDATE	LOS
702	080707	080921	77
702	611116	611129	14

N = 2

SUBJECT	READMIT	DSCHDATE	LOS
703	611116	620210	87
703	610515	610802	80

N= 2

SUBJECT	READMIT	DSCHDATE	LOS
704	611116	611121	6

n=1

Output 2: Correct Results

1. Report1: Study Subjects Only

SUBJECT	ADMDATE2	DSCHDATE	LOS
700	950313	950316	4
702	950315	950328	14
703	950315	950401	18
704	950315	950320	6
706	950316	950324	9
707	950316	950321	6
708	950317	950322	6
709	950317	950324	8
[...]			
727	950328	950402	6
728	950328	950404	6
[...]			

2. Report 2: Readmissions

SUBJECT	READMIT	DSCHDATE	LOS
708	950410	950415	6
708	950420	950423	64

N=2

SUBJECT	READMIT	DSCHDATE	LOS
709	950403	950428	6

N = 1

SUBJECT	READMIT	DSCHDATE	LOS
727	950426	950428	3

N= 1

SUBJECT	READMIT	DSCHDATE	LOS
728	950412	950414	3

N= 1