
A Technique for Handling Variable Length Character Variables in SAS

Rick Watts

Edmonton SAS Users Group

09 Jun 2005

Rick.Watts@globaliq.com

The Problem

- Clinical database needed to store study data in a consistent format regardless of source or data type.
- Designers chose Oracle CLOB datatype.
- Not currently supported by SAS
- Data translated to VARCHAR(2000)
- Supported by SAS/ACCESS

but....

The Problem

- SAS does not support variable length character variables.
- SAS/ACCESS assigns Varchar(2000) variables a length of 2000.
- Datasets large – may be gigabytes
- Inefficient to process and sort.

Example Oracle Table

Name	Null?	Type
-----	-----	-----
CRF_ID	NOT NULL	VARCHAR2(10)
PATIENT_ID	NOT NULL	VARCHAR2(20)
QUESTION1		VARCHAR2(2000)
QUESTION2		VARCHAR2(2000)
RECORD_ID	NOT NULL	NUMBER(10)

Resultant SAS Dataset

Alphabetic List of Variables and Attributes

#	Variable	Type	Len	Format	Informat	Label
2	CRF_ID	Char	10	\$10.	\$10.	CRF_ID
1	PATIENT_ID	Char	20	\$20.	\$20.	PATIENT_ID
3	QUESTION1	Char	2000	\$2000.	\$2000.	QUESTION1
4	QUESTION2	Char	2000	\$2000.	\$2000.	QUESTION2

57,000 observations occupy 229 Mb.

The Solution

- SAS not set to support variable length variables just yet.
- Need to ascertain required variable lengths programmatically.
- Create datasets accordingly.

Calculate Maximum Column Lengths

```
proc sql noprint;
  connect to oracle as ora (user=&oraUser
    password=&oraPass path =&oraPath);
  %put &sqlxmsg;
  select q1,q2 into :q1,:q2
  from connection to ora
  ( select      max(length(question1)) as q1 ,
              max(length(question2)) as q2
    from crf );
  disconnect from ora;
quit;
```

‘Fix’ The Macro Variables

- Remove unwanted spaces.
- Round up in anticipation of changes in data lengths.

```
%macro roundUp(var,percent);  
    %let &var =  
        %trim(%sysfunc(round(&&&var+&&&var/100*&percent,1)));  
%mend;
```

Create View of Oracle Table

```
proc sql;
connect to oracle as ora
    (user=&oraUser password=&oraPass path =&oraPath);
%put &sqlxmsg;
create view mydata as
select  patient_id, crf_id,
        question1 length = &q1 format = $&q1.. informat = $&q1..,
        question2 length = &q2 format = $&q2.. informat = $&q2..
from connection to ora
    ( select  patient_id, crf_id, question1, question2
      from    crf );
disconnect from ora;
quit;
```

<The Code>

Resultant View

Alphabetic List of Variables and Attributes

#	Variable	Type	Len	Format	Informat	Label
2	CRF_ID	Char	10	\$10.	\$10.	CRF_ID
1	PATIENT_ID	Char	20	\$20.	\$20.	PATIENT_ID
3	question1	Char	13	\$13.	\$13.	QUESTION1
4	question2	Char	32	\$32.	\$32.	QUESTION2

57,000 observations would occupy 4 Mb.

Since it's a view it actually occupies 5 K and the data is retrieved each time the view is accessed.

A Technique for Handling Variable Length Character Variables in SAS

Rick Watts

Edmonton SAS Users Group

09 Jun 2005

Rick.Watts@globaliq.com