SAS GLOBAL FORUM 2020 #SASGF

Paper 5114-2020

How a SAS Function and the ODS PACKAGE Statement Will Help You to Save Money and Space

Ricardo Rosales, Westat Inc.

ABSTRACT

In these times when data storage is costly, organizations often keep their data in compressed files on their servers to save both money and space.

It can be tedious and time consuming to take data out of compressed files, use the data, and then delete the extracted files. However, with the use of the GETOPTION(work) function, the contents of a dataset in a compressed file can be read and processed in the work library.

This paper demonstrates how to unzip files into the work library and how to create new zip files using the ODS Package Destination. This will keep your network space in a good shape.

INTRODUCTION

Zipping is a common method of archiving large files. The size of zipped files is considerably less than unzipped files, making them convenient for storage and file sharing, and when you email data in an attached zip file.

By keeping the data in zip files, projects can save a considerable amount of money. For example, if your organization charges a project $80 per gigabyte for storage and the size of an unzipped folder is 2.328G, the monthly charge will be $186.24. Compare this with the size after being zipped – 96M. The charges will then be $7.68, an almost 96% in savings by just keeping the data zipped.

When data is zipped, one of the challenges is that sometimes you will have a very large zip file that needs to be unzipped to use only a couple of files or datasets stored in it. Unzipping a large zip file can be frustrating. It takes time, patience, and a reliable computer. You may also need to close most of your other files and programs to make sure that your computer has low CPU usage and the space to unzip. This decreases the chance of your computer freezing or crashing.

When you email zip files, it reduces email transmission time and is convenient when you or your client have a mailbox size limit.

In the example in this paper, we will read a dataset from an existing zip file, load it into the work library, concatenate the extracted dataset for the year 2017 with a dataset for the year 2019, produce two report files, and put the report files in a zip file to email to a user. Finally, the datasets will be archived into a zip file for the year 2019, and then the dataset will be deleted from the folder. We break this task into 5 steps. The sections below describe each step in turn. The entire SAS program can be found in Appendix A.

STEP 1. READ THE DATASET FROM THE ZIP FILE.

To avoid opening the zip file we use the FILENAME zip method to read zip2017.zip into our SAS program. The zip method was added in SAS 9.4, so you can avoid using utilities such as WinZip, gunzip, or 7-Zip.

Using the FILENAME ZIP access method, we extract the 2017 dataset, electric2017.sas7bdat:

```
filename readzip zip "C:\temp1\zip2017.zip";
```

We unzip the ELECTRIC2017 dataset in the work library using GETOPTION(work) function:

```
filename ds "%sysfunc(getoption(work))/electric2017.sas7bdat" ;

data _null_;
  infile readzip(electric2017.sas7bdat)
  lrecl=256 recfm=F length=length eof=eof unbuf;
  file ds lrecl=256 recfm=N;
  input;
  put _infile_ $varying256. length;
  return;
  eof:
  stop;
run;
```

STEP 2. MERGING THE DATA AND PRODUCING THE OUTPUT.

Concatenate the electric dataset from 2017 and the electric dataset from 2019, then run PROC FREQ and PROC MEANS, to create two report files to send to users:

```
libname out19 "C:\temp1\data2019";

data years1719;
  set out19.electric2019
      work.electric2017;
run;

proc freq data=years1719;
  tables year * customer * revenue /
  out=freqElectric1719;
run;

PROC means DATA=years1719 maxdec=3;
  class year ;
  output out= meansElectric1719;
RUN;
```

STEP 3. CREATE A ZIP FILE FOR THE REPORTS.

Delete the datasets that we don't need anymore (electric2017 years1719) before zipping the files:

```
proc datasets library=work noprint;
  delete electric2017 years1719 ;
run;
```

Use the ODS OUTPUT and the PROC DATASETS procedure to create a dataset in the work library with the list of the datasets remaining in the work library (menasElectric1719 and meansElectric1719):

```
ods output Members=Members;
proc datasets library=work memtype=data;
run;
quit;
```

Next, create a zip file of all the datasets in the work library using the ODS PACKAGE destination. Make sure you use the NOPF option. The NOPF option tells SAS that the SAS

publishing framework is not needed. If you do not specify this option you may get this error:

```
ERROR: Not licensed for publish/subscribe access.
```

ODS package destination allows you to simply and easily create zip files from your ODS output:

```
ods package(newzip) open nopf ;
```

Add (adds a file or data set) and file= (specifies the file that you want to add) are required arguments when you are creating the zip file:

```
data _null_;
  set members;
call execute('ods package(newzip) add
file="%sysfunc(getoption(work))\'||strip(name)||'.sas7bdat";');
run;
```

Publish and archive builds the ODS package, provides a name for the zip file (archive_name=), and specifies the location to archive the file (archive_path=):

```
ods package(newzip) publish archive
  properties
 (archive_name="files4email.zip"
  archive_path="c:\temp1\" )
;
ods package(newzip) close;
```

STEP 4. SEND THE EMAIL WITH THE REPORTS.

Send an email with the zip file created in the previous step (files4mail.zip).

```
options emailsys=smtp emailhost='relay.westat.com' ;

FILENAME mailout EMAIL;

DATA _NULL_;

  FILE mailout;
  PUT '!EM_FROM!' 'ricardorosales@westat.com';
  PUT '!EM_to!' 'santaclaus@northpole.com';
  PUT '!EM_subject! Frequencies 2017 - 2019';
  PUT '!EM_ATTACH! C:\temp1\files4email.zip';
  PUT 'Attached files are frequencies for the years 2017 and 2019';

RUN;
```

STEP 5. ARCHIVE THE DATASETS TO A ZIP FILE.

Use the ODS PACKAGE destination to zip the entire contents of the folder for the year 2019. This creates a zip file at the location C:\temp1\data2019\zip2019.zip:

```
%macro drive(dir,zipname,ziploc);
  %let filerf=mydir;
  %let rc=%sysfunc(filename(filrf,&dir));
  %let did=%sysfunc(dopen(&filrf));
  %let memcnt=%sysfunc(dnum(&did));

  %let cnt=0;
```

```
%do i = 1 %to &memcnt;
%let ext=%upcase(%scan(%qsysfunc(dread(&did,&i)),2));
%let fname=%scan(%qsysfunc(dread(&did,&i)),1);

%if &ext=SAS7BDAT %then %do;
%let cnt=%eval(&cnt+1);
%let name&cnt=&fname..&ext;
%end;

%end;
ods package open nopf;

%do j = 1 %to &cnt;

ods package add file="&dir.&&name&j" mimetype="application/x-compress";
ods package publish archive properties(archive_name="&zipname"
 archive_path="&ziploc");
%end;

ods package close;

%let rc=%sysfunc(dclose(&did));
```

After the zip file is created, use PROC DATASETS, with the kill option to delete all the datasets in the folder:

```
libname deldata "&dir";

proc datasets library=deldata kill;
run;
quit;
%mend drive;
%drive (C:\temp1\data2019\,zip2019.zip,C:\temp1\);
```

CONCLUSION

Archiving data in zip files reduces the amount of space used on your network and reduces costs. You can open datasets from a zip file in the work library without unzipping the entire zip file. You can create a new zip file with the ODS package. The limitation is that you cannot add files to an existing zip file. Emailing zip files reduces the transmission time.

REFERENCES

Hemedinger, Chris. “Using FILENAME ZIP to unzip and read data files in SAS.” *SAS Blogs*. May 11, 2015. Available at https://blogs.sas.com/content/sasdummy/2015/05/11/using-filename-zip-to-unzip-and-read-data-files-in-sas/.

Hamilton, Jack. “Creating Zip Files with ODS.” Division of Research, Kaiser Permanente, Oakland, California. April 1, 2013. Available at http://support.sas.com/resources/papers/proceedings13/131-2013.pdf

DISCLAIMERS

The contents of this paper are the work of the author and do not necessarily represent the opinions, recommendations, or practices of Westat.

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

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

ACKNOWLEDGMENTS

The author would like to thank Westat management for supporting his participation in SAS Global Forum 2020, and Michael Raithel for help in reviewing his paper. A special thank you for my wife Laurie for all her support.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Ricardo Rosales
Westat
(301) 517-8048
RicardoRosales@westat.com

APPENDIX A

```
/*=================================================================*/;
/*Step 1.Read the dataset from the zip file*/;
/*=================================================================*/;
filename readzip ZIP "C:\temp1\zip2017.zip";

/* Copy a zipped data set into the WORK library Using the System function
    Getoption(work) */

filename ds "%sysfunc(getoption(work))/electric2017.SAS7BDAT" ;

data _null_;
   infile readzip(electric2017.SAS7BDAT)
      lrecl=256 recfm=F length=length eof=eof unbuf;
   file   ds lrecl=256 recfm=N;
   input;
   put _infile_ $varying256. length;
   return;
eof:
   stop;
run;

/*=================================================================*/;
/*Step 2. Merging the data and producing the output*/;
/*=================================================================*/;
/* get data from 2019 to generate frequency for 2017 X 2019 */;

libname out19 "C:\temp1\data2019";
```

```
data years1719;
      set out19.electric2019 work.electric2017;
run;

proc freq data=years1719;
      tables year * customer  * revenue /
      out=freqElectric1719;
run;

PROC means DATA=years1719 maxdec=3;
      class year ;
      output out=meansElectric1719;
RUN;

proc datasets library=work noprint;
   delete electric2017 years1719 ;
run;


/*=========================================================================*/;
/*Step 3. Create a Zip file for the reports                                */;
/*=========================================================================*/;
/* creates a dataset with a list of the datasets of the work library */;

ods output Members=Members;
proc datasets library=work memtype=data;
run;
quit;

/* Creating a ZIP file with the frequencies with ODS PACKAGE */
ods package(newzip) open nopf ;

/* reads the members dataset to get the names of the dataset of the work
library and add them to the zip file */;
data _null_;
      set members;
      call execute('ods package(newzip) add
file="%sysfunc(getoption(work))\'||strip(name)||'.sas7bdat";');
run;

ods package(newzip) publish archive
  properties(archive_name="files4email.zip" archive_path="c:\temp1\" );
ods package(newzip) close;

/*=========================================================================*/;
/*Step 4. Send the email with the reports*/;
/*=========================================================================*/;
/* Email frequencies in the zip file */;

options emailsys=smtp emailhost='relay.westat.com' ;

FILENAME mailout EMAIL;

DATA _NULL_;

  FILE mailout;
```

```
    PUT '!EM_FROM!' 'ricardorosales@westat.com';
    PUT '!EM_to!' 'SASHelp@westat.com';
    PUT '!EM_subject! Frequencies 2017 - 2019';
    PUT '!EM_ATTACH! C:\temp1\files4email.zip';
    PUT  'Attached files are frequencies for the years 2017 and 2019';

  RUN;

/*=====================================================================*/
  /*Step 5. Archive the datasets for the year 2019 in a zip file*/;
  /*=====================================================================*/;
  /* macro to create a zip file with all datasets from a directory*/;

  %macro drive(dir,zipname,ziploc);
     %let filerf=mydir;
     %let rc=%sysfunc(filename(filrf,&dir));
     %let did=%sysfunc(dopen(&filrf));
     %let memcnt=%sysfunc(dnum(&did));

     %let cnt=0;
     %do i = 1 %to &memcnt;
        %let ext=%upcase(%scan(%qsysfunc(dread(&did,&i)),2));
        %let fname=%scan(%qsysfunc(dread(&did,&i)),1);

        %if &ext=SAS7BDAT %then %do;
           %let cnt=%eval(&cnt+1);
           %let name&cnt=&fname..&ext;
        %end;

     %end;
       ods package open nopf;

     %do j = 1 %to &cnt;

        ods package add file="&dir.&&name&j" mimetype="application/x-
  compress";
        ods package publish archive properties(archive_name="&zipname"
  archive_path="&ziploc");
     %end;

     ods package close;

     %let rc=%sysfunc(dclose(&did));

  /* Remove all datasets from the folder */

  libname deldata "&dir";

  proc datasets library=deldata kill;
  run;
  quit;
  %mend drive;
  %drive (C:\temp1\data2019\,zip2019.zip,C:\temp1\);
```