SAS® GLOBAL FORUM 2019

USERS PROGRAM

APRIL 28 – MAY 1, 2019 | DALLAS, TX

PERFORMRx℠

Next Generation Pharmacy Benefits

Downloading zipped files and Extracting data from External FTP sites using SAS

Shimels Afework, Ricardo Alvarez, and Sammi Nguyen

PerformRx

Shimels Afework

Abstract

Accessing external ftp sites to download files is a resource consuming manual process. This inefficiency increases more when the files are large, many, and zipped.

The common practice to acquiring these data from external ftp sites is to go to the secured site, log in using user id and password, find the location of the file with in the directory structure, download the files to the target location, and unzip each file to extract the data.

The speed at which this repetitive process can be completed depends on network traffic, distractions during this process and other similar factors.

This paper provides a SAS based automated process which can be scheduled on a windows or Linux environment to complete the entire downloading and extraction process at any given day and time.

PERFORMRx℠

Next Generation Pharmacy Benefits

Downloading zipped files and Extracting data from External FTP sites using SAS

Shimels Afework, Ricardo Alvarez, and Sammi Nguyen

PerformRx

Introduction

As organizations' grow, productivity becomes more and more critical. Automating repetitive tasks improves productivity and reduces cost. In addition to consuming significant resources, these tasks can also get skipped from time to time because of overlapping responsibilities and the hard decision to choose between competing tasks.

The SAS scheduler is a very useful tool to automatically execute tasks by scheduling jobs at a convenient day and time. In this paper, we will discuss the substantial efficiency we have gained by scheduling a repetitive task of downloading zipped files from external FTP site and extracting the data using SAS scheduler.

PERFORMRx℠

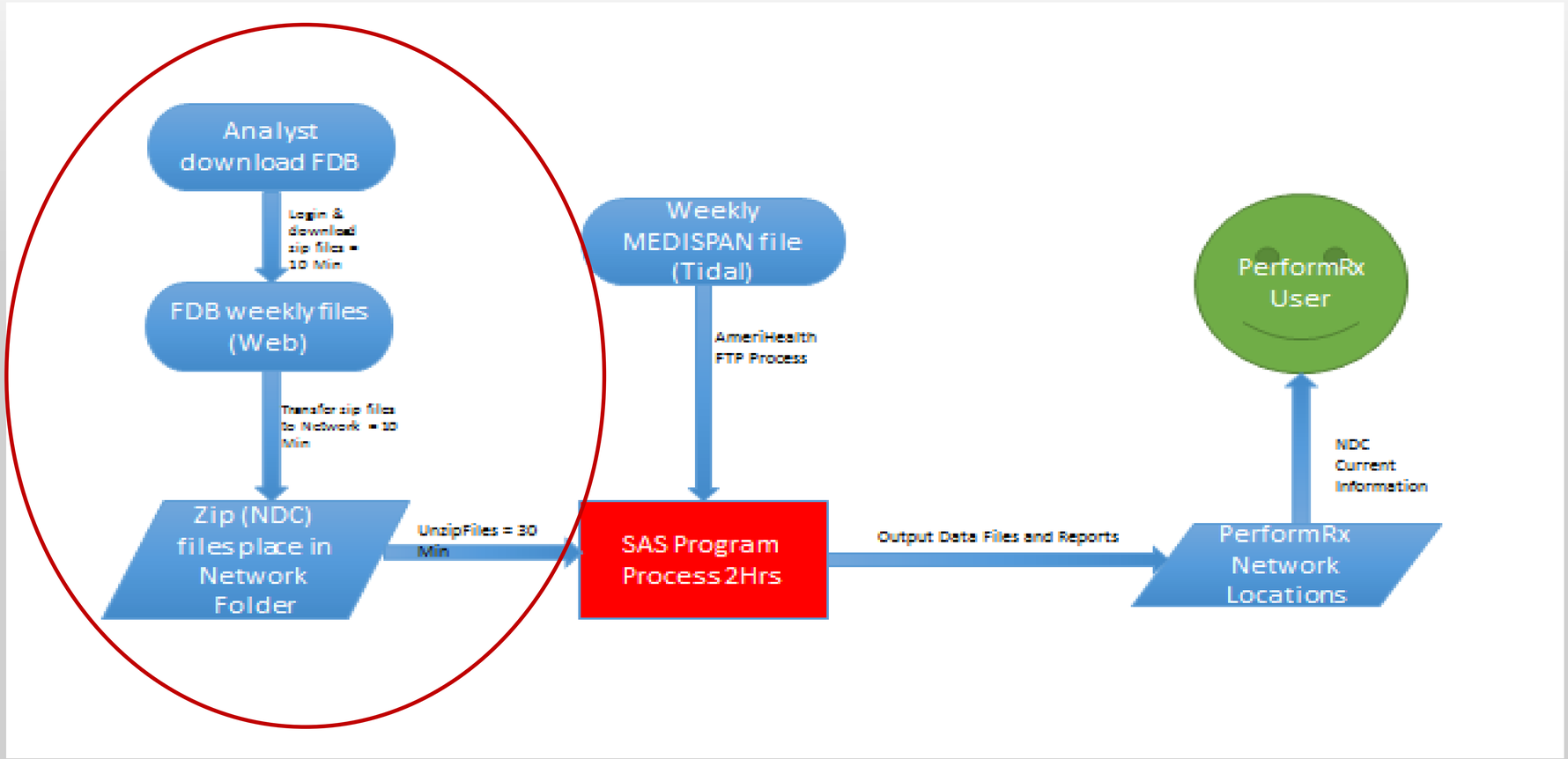
Next Generation Pharmacy Benefits

Downloading zipped files and Extracting data from External FTP sites using SAS

Shimels Afework, Ricardo Alvarez, and Sammi Nguyen

PerformRx

Manual Process Work Flow



PERFORMRx℠

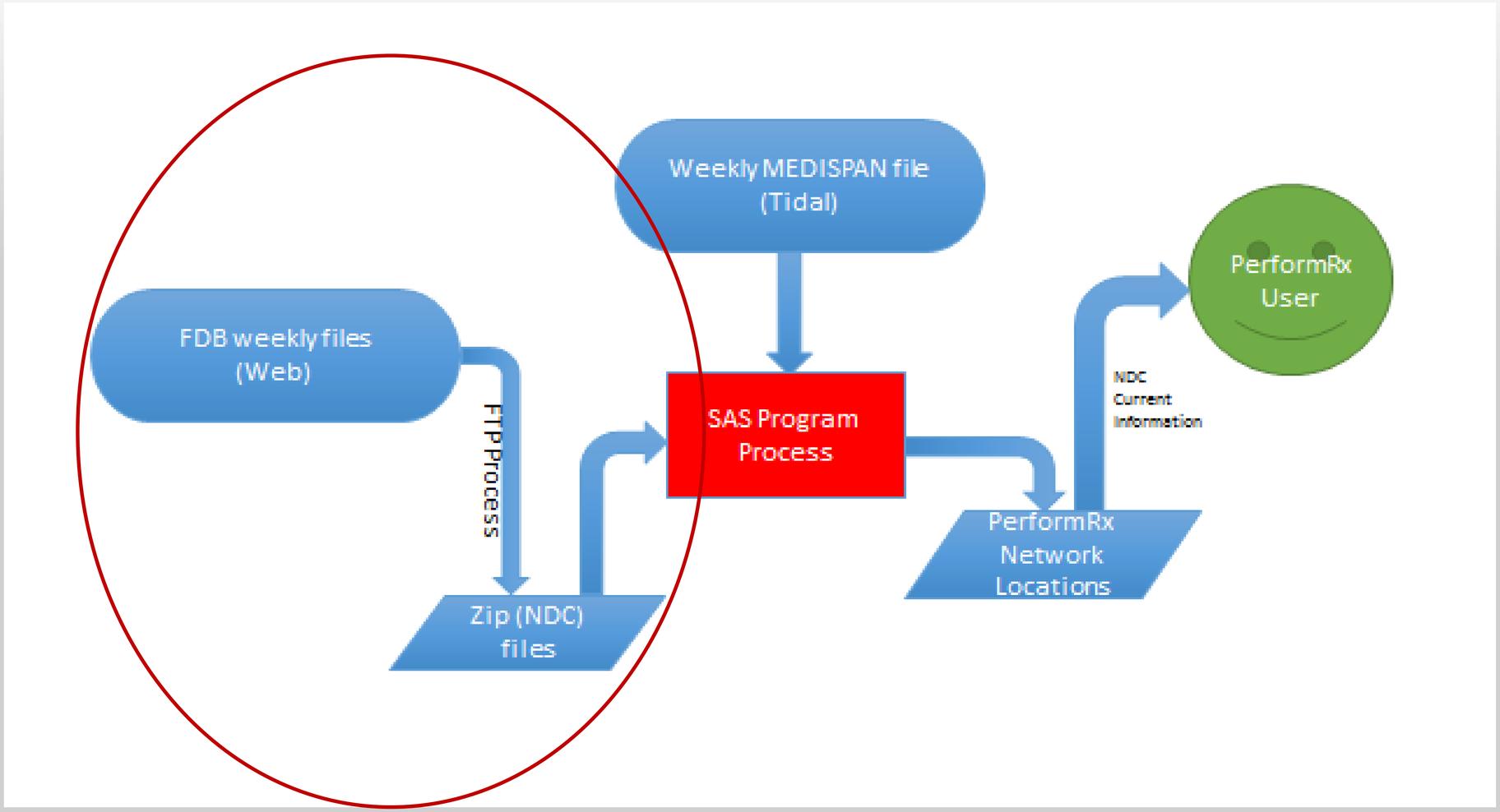
Next Generation Pharmacy Benefits

Downloading zipped files and Extracting data from External FTP sites using SAS

Shimels Afework, Ricardo Alvarez, and Sammi Nguyen

PerformRx

SAS Automated Process Work Flow



PERFORMRx℠

Next Generation Pharmacy Benefits

Downloading zipped files and Extracting data from External FTP sites using SAS

Shimels Afework, Ricardo Alvarez, and Sammi Nguyen

PerformRx

SAS Program Code

Setting up system/file date

```
%LET FILEDATE =  %SYSFUNC(PUTN(%SYSEVALF(%SYSFUNC(TODAY())-      1),YYMMDDN8.));
%PUT &FILEDATE.;
```

Setting up Linux library and location

```
LIBNAME LIB_NM "/DIRECTORY_NAME/SUB_ DIRECTORY_NAME "; /*Library*/
%LET LIBNM_OUT"/DIRECTORY_NAME/SUB_DIRECTORY_NAME/XX.ZIP; /*location of the file*/
```

Clearance to LINK to source file

```
FILENAME FILENM_1 /* the file name that will be calling */              USER='XXXXXX' /****  required field
*/
PASS='XXXXXX' /****  required field  */
HOST='FTP. SITE_NAME.COM'/*the url address to access the files*/
CD="/DIRECTORY_NAME/SUB_ DIRECTORY_NAME /" /*subdirectories on url*/
RECFM=S DEBUG /* specifies the record format of the external file; */
FTP 'FILE_2.ZIP'; /*  the zip file name on the FDB server  */
```

PERFORMRx℠

Next Generation Pharmacy Benefits

Downloading zipped files and Extracting data from External FTP sites using SAS

Shimels Afework, Ricardo Alvarez, and Sammi Nguyen

PerformRx

SAS Program Code

Creating a placeholder for the DATA Set in the zip files

```
DATA _NULL_;
  INFILE FILENM_1  NBYTE=n;
  FILE "&FILENM." recfm=n ;
  N=1;
  INPUT;
  PUT _INFILE_ @@;
run;

filename ZIP_FILENM ZIP "&FILENM."; /*calling the location with a filename */
```

Placing the zip files in Linux folder and Reading the DATA Set

```
DATA FLSNM(KEEP=MEMNAME); /*read individual files from the ZIP file */
      FILENAME  ZIP_FILENM ZIP "&FILENM.";
      LENGTH MEMNAME $200;
      FID=DOPEN("ZIP_FILENM ");
      IF FID=0 THEN
          STOP;
      MEMCOUNT=DNUM(FID);
      DO I=1 TO MEMCOUNT;
          MEMNAME=DREAD(FID,I);
          OUTPUT;
      END;
      RC=DCLOSE(FID);
RUN;
```

PERFORMRx℠

Next Generation Pharmacy Benefits

Downloading zipped files and Extracting data from External FTP sites using SAS

Shimels Afework, Ricardo Alvarez, and Sammi Nguyen

PerformRx

SAS Program Code

Generate a report of all the FILES contained in the zip file

```
TITLE "FILES IN THE  FLSNM ZIP FILE";
PROC PRINT DATA= FLSNM NOOBS N;
```

memname
NDDF Plus DB/Counseling Messages/CMM 1.0/RCMMD0_DESC
NDDF Plus DB/Counseling Messages/CMM 1.0/RCMMGC0_GCNSEQNO_LINK
NDDF Plus DB/Dosage Range Check/DRCM 3.0/RDRCAT0_ADJ_TYPE
NDDF Plus DB/Dosage Range Check/DRCM 3.0/RDRCCD0_CALC_REQ_TYPE_DESC
NDDF Plus DB/Dosage Range Check/DRCM 3.0/RDRCCVU0_UNITS_CONVERSION
NDDF Plus DB/Dosage Range Check/DRCM 3.0/RDRCDD0_DOSE_CALC_DESC
NDDF Plus DB/Dosage Range Check/DRCM 3.0/RDRCDTD0_DOSE_TYPE_DESC
NDDF Plus DB/Dosage Range Check/DRCM 3.0/RDRCEX0_EXCLUSIONS
NDDF Plus DB/Dosage Range Check/DRCM 3.0/RDRCMA1_MSTR
NDDF Plus DB/Dosage Range Check/DRCM 3.0/RDRCMD0_MATH_PROCESS_DESC

PERFORMRx℠

Next Generation Pharmacy Benefits

Downloading zipped files and Extracting data from External FTP sites using SAS

Shimels Afework, Ricardo Alvarez, and Sammi Nguyen

PerformRx

Results

The SAS automated file downloading and data extraction process increased our overall operational efficiency by 50%.

Companies depend on various vendor databases that are required to be downloaded and extracted before use. Depending on the frequency to perform this task and the size of the external databases to be downloaded and extracted; the 50% operational efficiency can result in significant amount of resource and cost savings.

PerformRx is now expanding the implementation of this automated process to other vendor products.

Downloading zipped files and Extracting data from External FTP sites using SAS

Shimels Afework, Ricardo Alvarez, and Sammi Nguyen
PerformRx

Conclusion

Process automation improves efficiency by significantly reducing time and other resources needed to complete manual tasks. It also allows saved time and resource to be utilized to accomplish more important tasks.

As we have discussed in this paper, by utilizing the SAS scheduler, PerformRx was able to improve processing time of downloading zipped files and extracting data from external FTP sites by one-half, thus increasing productivity by 50%.

References

AlZahal, O., B. Rustomo, N. E. Odongo, T. F. Duffield, and B. W. McBride. 2007. Technical note: A system for continuous recording of ruminal pH in cattle. J Anim Sci 85(1):213-217.

Cassell, D. L. 2007. Don't Be Loopy: Re-Sampling and Simulation the SAS Way. in SAS Global Forum 2007 Conference. S. I. Inc, ed. SAS Institute Inc.

Enemark, J. M. 2008. The monitoring, prevention and treatment of sub-acute ruminal acidosis (SARA): a review. Vet J 176(1):32-43.

Sato, S., A. Ikeda, Y. Tsuchiya, K. Ikuta, I. Murayama, M. Kanehira, K. Okada, and H. Mizuguchi. 2012. Diagnosis of subacute ruminal acidosis (SARA) by continuous reticular pH measurements in cows. Vet Res Commun 36(3):201-205.

Zhang, Y. and Y. Yang. 2015. Cross-validation for selecting a model selection procedure. Journal of Econometrics 187(1):95-112.