

## Paper 3100 - 2015

### Using SAS® to Manage SAS® Users on UNIX File System

Venkateswarlu Toluchuri, United Health Group, Hyderabad, India.

#### ABSTRACT

SAS platform administrators always feel the pinch of not having information on how much storage space is occupied by each user on one specific file system or in the entire environment. Sometimes the platform administrator does not have an access to all users' folders. So they have to plan for the worst. There are multiple approaches to tackle this problem. One of the better methods is to initiate an alert mechanism to notify a user when they are in the top 10 file system users on the system.

#### INTRODUCTION

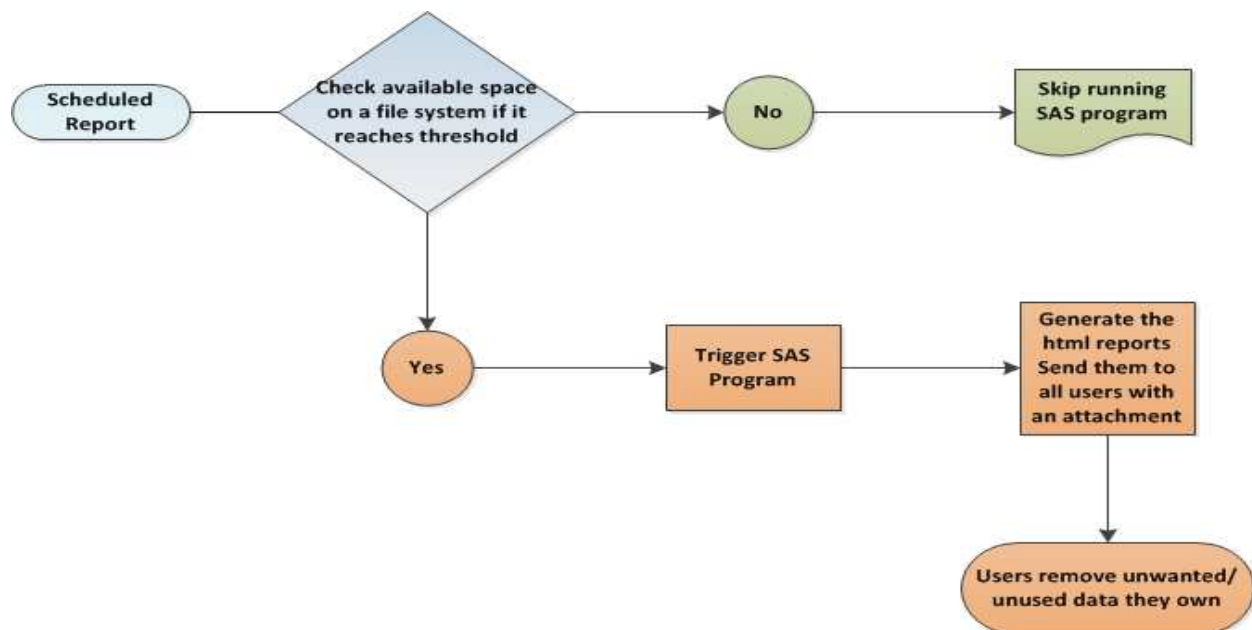
This paper talks about the approach I am proposing to be used in our AIX environment to keep track of storage space usage by each user and alerting the users via email. The process uses SAS® data step functions for reading metadata to obtain email addresses for emailing notification to the users.

#### APPROACH

All production analytics AIX platforms contain multiple file system areas to accommodate never-ending demand from SAS analyst and programmer user communities. We have many business file systems deployed and one user may belong to more than one business cluster having the ability to write data to multiple file systems. Usually business analyst users or programmers create huge permanent datasets for their analysis and they typically forget to clean up those files after they are no longer needed. Our file systems are always approaching total use of the available space. This results in frustration for users trying to create more files for current projects and administrators trying to determine who has filled up storage space and how much. Multiple techniques were used in the past including providing guidance to users, defining standards of programming and mandating users to follow them, etc. I have devised a three-step preventive mechanism to address this issue.

In order to run 'x' command in the script, you will need to get this run with root account or ask them to grant you the ability to run it with your UNIX account. 'x' command generates a file containing a list of UNIX accounts, file names, and dates when the file has been updated or created. A SAS program with a SAS Data step creates a SAS Dataset with summarized data from the file and lists the top users alerts them via email communication with desired report.

#### Flow diagram for users level space report:



## SAS Program

**First step: Find current available space on a File System, if it breaches to threshold point (considered here is 85%) it triggers sas program to capture all users file content.**

```
%macro filesysize;
    filename pro11 pipe " df -g /<filesystem> |tail -1 "; /***** captures available space percentage on a file
system*****/;
    *** parse information from unix ***;
    data usr;
        length filesys $20 gb_blocks 8 free_space 8 persen_used $8 iused 8 persen_iused $8 mount $20;

        ***** read standard output from unix command *****;
        infile pro11 lrecl=162 missover pad;
        input filesys gb_blocks free_space persen_used iused persen_iused mount;

    run;

    data usr1;
        set usr;
        date_time= put(datetime(),datetime18.);
        persen_used1 = input(substr(persen_used,1,length(persen_used)-1),8.);

    run;

    proc sql noprint;
        select persen_used1 into :persen from usr1;
    quit;

    %if &persen > 85 %then
        %do;
            x '/XXXXXXXX/XXXXXXXX/every4hour/unix_user_report.sas &';
        %end;
%mend filesysize;
%filesysize;
```

**Second Step: create data step and generate the final report.**

```
/* unix_user_report.sas start from here */
```

```
/* Get sudo access on find command or run with root to avoid permission issues on user folders and here
considered to get all files size more than 1 GB */
```

```
x "find /<filesystem> -xdev -size +1024 -exec ls -l {} ';' > /tmp/test.txt";
```

```
data processes (drop=__ : fper t filesize t);
    format UID $20. Filesize Location $500. Gb;
    infile "/tmp/XXXXXXXX" lrecl=2000 truncover firstobs=1;
    input;
    __line=__infile__;
    Fper=compress(scan(__line,1,' '),');
    t=compress(scan(__line,2,' '),');
    UID=compress(scan(__line,3,' '),');
    Group=compress(scan(__line,4,' '),');
    Filesize=compress(scan(__line,5,' '),');
    Month=compress(scan(__line,6,' '),');
    Date = compress(scan(__line,7,' '),');
    Time=compress(scan(__line,8,' '),');
    Location=compress(scan(__line,9,' '),');
    GB=(filesize)/(1024*1024*1024);
```

```

        File_Name=scan(location,-1,'/');
run;

%macro create(UID);

ods html body="/aalsasmon_apm/Reports/&UID..html" style=sasweb;
proc print data=processes (keep=UID GB Location File_Name Month Date Time) noobs;
    where UID="&UID";
    title "Report for a &UID on /aalpprg filesystem ";
run;
ods html close;

%mend;

proc sort data=processes out=newdsn nodupkey;
by UID;
data _null_ ;
    set newdsn;
    by UID;
    call execute('%create('||UID||')');
run;

data one;
    set newdsn;
    retain fmtname '$test';
    rename UID=start ;
    label='<A '||compress('HREF="||UID||'.html')
        ||">||trim(UID)||'</A>';
run;

proc format cntlin=one;
run;

proc sql;
    create table users as select distinct(UID) from processes;
    select count(*) as obscnt into :cnt from users;
    select UID into: uid1 separated by " " from users;
quit;

/* Affected users are informed via Email. Simple data step iteration is used to send emails to individual users. Below
code uses Email Access Method */
libname input "/XXXXXX/monitoring_data/metadata_xtract";

data email_list;
    set input.email(keep = keyid emailaddr externalkey );

    if externalkey = '0' then
        delete;
    uid=compress (keyid);
run;

proc sql;
    create table email_add as select a.*,b.Name from email_list a, input.email_info b
        where a.keyid=b.Person_ExtId_Identifier;
run;

data email_add;
    set email_add;
    Name=scan(Name,1,'-');
run;

proc sql;

```

```

        create table user_report as select a.*,email_add.Name from processes a, email_add where
a.uid=email_add.uid;
run;

data user_report;
set user_report;
if Name="" then Name='ORPHAN';
if emailAddr="" then emailAddr='No Email Found';
run;

options symbolgen mlogic mprint;

proc sql;
    create table email as select * from email_list where keyid in("&uid1");
    select emailaddr into: email1 separated by "" from email;
quit;

%let timenow=%sysfunc(time(), time.);
%let datenow=%sysfunc(date(), date9.);

goptions hsize=5in vsize=5in;
title1 "PLEASE FIND THE UTILIZATION REPORT FOR /aalpprg ON APSPT0152 ON &datenow. &timenow";

ods tagsets.htmlpanel file="/xxxxxxx /xxxxx/test_filesystem.html" style=sasweb event=panel(start);
proc gchart data=user_report
;
    hbar3d
    UID
/
    sumvar=gb
    shape=block
frame discrete
    type=sum
sum
    legend=legend1
    descending
    coutline=black

patternid=midpoint
;

run;
ods listing close;
ods html file='XXXXXXXXXX/XXXXX.html' style=sasweb;
title "please find the user utilization report for xxx file system on prod environment on &datenow. &timenow";

/* Generate the html report using processes dataset */
proc report data=user_report nowd;
column UID gb Name emailAddr ;
define UID / group 'UID' ;
define GB /analysis sum format=8.2;
define name /group 'Name';
define emailAddr / group 'emailAddr';
compute UID ;
    href=trim(UID)||".html";
    call define(_col_, "URLP", href);
endcomp;
rbreak after / summarize;
run;

ods tagsets.htmlpanel event=panel(finish);

```

```
ods tagsets.htmlpanel close;
ods html close;
```

/\*Affected users are informed via Email. Simple data step iteration is used to send emails to individual users. Below code uses Email Access Method \*/

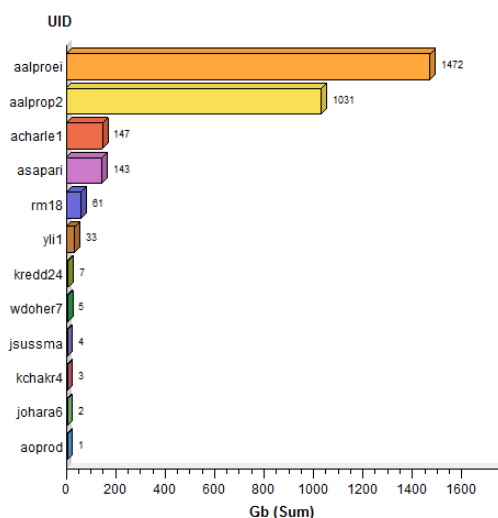
```
filename outbox email "venkat.toluchuri@example.com";
```

```
data _null_;
    file outbox
        to("&email1")

    /* filename statement */
    cc=("xxxxxxx@xxx.com" )
    subject = "urgent :- filesystem : <filesystemname> is filled more than 80 percent. Please do cleanup owned
by your files if no more used"
    attach="/xxxxxxx/<filesystem>.html";
    put " ";
    put "Hello,";
    put " ";
    put " ";
    put "Disk Utilization Report for <filesystem> file systme on PROD by user level";
    put " ";
    put " ";
    put "This Email was sent from an auto-notification system. Please do not reply to the above email address.";
    put " ";
    put " ";
    put " ";
    put " ";
    put " Thanks";
    put " ";
    put " ";
    put " ";
run;
```

### Sample Output for summarized report for a file system:

PLEASE FIND THE SAS USERS UTILIZATION REPORT FOR /TEST FILE SYSTEM ON PROD ENVIRONMENT ON 17DEC2014 6:25:10



Obs	UID	Gb	Name	emailAddr
1	aalproei	1471.75	Stephen	user.report@sample.com
2	aalprop2	1030.78	Steve	user.report@sample.com
3	acharle1	146.84	Robert	user.report@sample.com
4	aoprod	1.02	Justin	user.report@sample.com
5	asapari	143.22	Alfred	user.report@sample.com
6	johara6	2.03	Jyothi	user.report@sample.com
7	jsussma	3.67	Ram	user.report@sample.com
8	kchakr4	3.16	Kavitha	user.report@sample.com
9	kredd24	7.02	Raju	user.report@sample.com
10	rm18	39.71	MATLA RADHAKRISHNA	user.report@sample.com
11	wdoher7	4.57	Bill	user.report@sample.com
12	yli1	33.05	Yim	user.report@sample.com
13	Grand Total	2907.72		

#### **Detailed output for one user report from above summarized report:**

UID	Name	Gb	Location	File_Name	Month	Date	Time
rm18	MATLA RADHAKRISHNA	24.00	/xxxxxxxx/analysis/rm18/xxx/rhonda/a.sas7bdat	a.sas7bdat	Aug	28	07:59
		4.07	/xxx/analysis/rm18/b.sas7bdat	b.sas7bdat	Aug	14	05:40
		3.31	/xxxxxxxx/xxxxxxxx/test.sas7bdat	test.sas7bdat	Jul	29	05:17
		3.08	xxxxx/xxxx/xxxx/xxxx/hello.sas7bdat	hello.sas7bdat	May	07	08:10
		1.75	xxxxx/xxxxxx/xxxxxx/xxxxxx/parent.sas7bdat	parent.sas7bdat	Sep	12	08:17
		1.75	xxxxx/xxxxxx/xxxxxx/xxxxxx/xxxxx/thanks.sas7bdat	thanks.sas7bdat	Sep	12	08:17
		1.75	xxxxx/xxxxxx/xxxxxx/xxxxxx/bc.sas7bdat	bc.sas7bdat	Sep	12	08:16
rm18		39.71					
Grand Total		39.71					

#### **CONCLUSION**

It is never easy to tell a user how much data his/her code created in a file system. So it is better to alert them and let them decide if they are really creating the desired output from the code or are they not being good stewards of the shared space they are using.

#### **REFERENCES**

1. <http://support.sas.com/documentation/cdl/en/lrdict/64316/HTML/default/viewer.htm#a002058232.htm>
2. <http://support.sas.com/kb/23/379.html>

#### **ACKNOWLEDGMENTS**

I would like to thank Michael A Criswell for his constant support and encouragement. I will also like to thank my team Chris D Schwab, Venkatasubbared Kottapalli and Jennings Whit Box for their valuable input and support.

#### **CONTACT INFORMATION**

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

Name: Venkateswarlu Toluchuri  
Enterprise: United Health Group  
Address: Hi-Tech City 2 Building H09, Hyderabad, India, 500081  
Work phone: +91-40-3085-2768  
E-mail: venkateswarlu.t@optum.com, venkateswarlutoluchuri@gmail.com

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.