

An Effective CLEANWORK.SAS[®] for Windows

John M. Wildenthal, JPMorgan Chase

ABSTRACT

SAS provides an effective utility for deleting orphaned work directories on Unix boxes. But their offering for Windows merely deletes everything older than a particular time (<http://support.sas.com/techsup/unotes/SN/008/008786.html>). This is unworkable for a server environment. This paper discusses a short program for Windows that parses the root WORK directory and matches directory names to SAS v9 running processes. Directories that are not matched are then deleted.

MOTIVATION

In a server environment, it is not unusual for some SAS sessions to terminate abnormally, leaving that process' work directory on the server. Over time, the number of these directories and the space they take can grow to be a problem.

SAS provides a binary for Unix that appears to match work directories to running processes, and then delete work directories that are not part of a running process.

SAS provides a SAS program for Windows. But this program merely tries to delete everything older than a particular time. That could mean your (running) metadata server's work directory, too.

This paper presents a SAS program I wrote to clean work directories on Windows boxes that do not belong to any running process. This avoids deleting data from your metadata server's work directory, and other directories associated with SYSTEM processes.

ELEMENTS

We need two things:

1. the list of active SAS processes
2. the list of work directories

SAS has no native way to access the list of active process Ids, henceforth PIDs. There are commands in Unix, which their provided binary may use, but there is no single command for Windows. The Windows Resource Kit for NT and W2K servers contains a program – tlist.exe – which provide PIDs. XP and W2K3 Server have tasklist.exe to provide equivalent functionality. For other Windows versions, or for systems that do not have the Windows Resource Kit, one can download a freeware program – Process Viewer from www.prcview.com – which will provide the necessary functionality.

Once we have the list of active running processes, how do we match directories?

There are several ways to determine the root of the work directories. I prefer to use SAS' PATHNAME function. Giving this function WORK as the argument returns the current WORK directory. I can then trim off the last directory to get the root of all the work directories.

To get the list of all work directories, we can use FILENAME PIPE with the DIR command.

It is apparent that SAS creates the normal work directory in Windows by appending the PID to “_TD”. But there is another kind of directory, a SAS_util directory. After doing a little research, you will find a hexadecimal version of the PIDs in those directory names.

So we have PIDs, and a list of directories with the PID that spawned them. Now we merge the two on PID and delete any found in the first list. The remainder is the list of directories to delete. Use a DATA _NULL_ and drop them into a .BAT file, CALL SYSTEM that bat file, and then you are done.

NOTE: This code may not work in 9.1.3 unless you have applied Service Pack 4

CODE:

```
/******  
** Tired of work directories from killed SAS jobs clogging your Windows **  
** box? Wish you had the equivalent of the "cleanwork" utility SAS **  
** provides for Unix? **  
** **  
** This program implements the equivalent of the unix version of **  
** cleanwork as a SAS program. The SAS provided program (SAS Note **  
** SN-008786) merely deletes everything older than a preset time. This **  
** code gets a list of running processes and only deletes the work **  
** directories that belong to processes that are no longer running. **  
** **  
** The main difficulty is getting a list of running tasks. Different **  
** versions of Windows have different tools available or none available. **  
** For those without any available, there are freeware tools available for **  
** this purpose. **  
** **  
** The choices are "tlist.exe" which is part of the Windows Resource Kit **  
** for NT and W2K servers, "tasklist.exe" for XP, and a freeware program. **  
** I do not know what might be available on Win95, 98, 98SE, ME, so for **  
** those I will recommend the use of the freeware program. The program is **  
** "Process Viewer", specifically the command line version. The code **  
** below was tested with version 5.2.1.2 of "pv.exe" available at **  
** http://www.prcview.com/ **  
** **  
** To use, uncomment the relevant section and verify the presence of the **  
** task listing program. You may need to change the path to match your **  
** configuration. After you are pointing at the correct tool, this job **  
** can be run in batch as a Scheduled Task to regularly clean your SAS **  
** Temporary folders of orphaned work directories. **  
** **  
** 01FEB2006 John M. Wildenthal **  
*****/  
%* Get list of running SAS jobs;  
  
/****** USING TLIST.EXE *****  
FILENAME tasklist PIPE "c:\progra~1\suppor~1\tlist.exe" ;  
  
DATA tasklist;  
  LENGTH pid 8 process_name $40;  
  INFILE tasklist;  
  INPUT;  
  pid = INPUT(SCAN(_INFILE_,1),8.);  
  process_name = UPCASE(SCAN(_INFILE_,2,' '));  
  IF INDEX(process_name,'SAS.EXE');  
RUN;  
*****/  
  
%***** USING TASKLIST.EXE *****;  
FILENAME tasklist PIPE '"c:\windows\system32\tasklist.exe"';  
  
DATA tasklist;  
  LENGTH pid 8 process_name $40;  
  INFILE tasklist FIRSTOBS=4;  
  INPUT;  
  pid = INPUT(SUBSTR(_INFILE_,27,8),8.);  
  process_name = UPCASE(SUBSTR(_INFILE_,1,25));
```

```

    IF INDEX(process_name, 'SAS.EXE');
RUN;
%*****;

/***** USING PV.EXE *****/
FILENAME tasklist PIPE "w:\sas\scheduled_builds\pv.exe";

DATA tasklist;
    LENGTH pid 8 process_name $40;
    INFILE tasklist;
    INPUT;
    pid = INPUT(SCAN(_INFILE_, 2, ' '), 8.);
    process_name = UPCASE(SCAN(_INFILE_, 1, ' '));
    IF INDEX(process_name, 'SAS.EXE');
RUN;
%*****/

%* Get list of all work directories - active and orphaned;

%LET workdir = %SYSFUNC(PATHNAME(work,L));
%LET rworkdir = %SYSFUNC(REVERSE(&workdir));
%LET rootwork = %SYSFUNC(REVERSE(%SUBSTR(%STR(&rworkdir), %INDEX(%STR(&rworkdir), \)+1)));
%LET dircmd = dir /ad %STR("%)&rootwork%STR("%);

FILENAME dirlist PIPE "&dircmd";

DATA workdirs;
    LENGTH dirname $80;
    INFILE dirlist;
    INPUT;
    dirname = _INFILE_;
    IF (SUBSTR(dirname, 1, 1) NE ' ') %* an actual entry, not header or footer;
    AND
    (SCAN(dirname, -1, ' ') NE '.') %* not the self directory;
    AND
    (SCAN(dirname, -1, ' ') NE '..'); %* not the parent directory;
    dirname = SCAN(dirname, -1, ' ');
    IF SUBSTR(dirname, 1, 1) EQ '_' THEN %* _TD form of directory name;
    pid = INPUT(SUBSTR(dirname, 4), 8.);
    ELSE %* util form, avoid machine name;
    pid = INPUT(REVERSE(SUBSTR(REVERSE(SCAN(dirname, 2, '_')), 1, 4)), HEX4.);
RUN;

PROC SORT DATA=workdirs;
    BY pid;
RUN;

PROC SORT DATA=tasklist;
    BY pid;
RUN;

%* orphaned work directories are not associated with a running job;

DATA orphandirs;
    MERGE workdirs (IN=dir_there)
          tasklist (IN=running_job);
    BY pid;

```

```

    IF dir_there AND NOT running_job;
RUN;

PROC CONTENTS DATA=orphandirs OUT=do_or_not NOPRINT;
RUN;

PROC SQL STIMER NOPRINT;
    SELECT DISTINCT nob
        INTO :do_or_not_var
        FROM do_or_not;
QUIT;

%MACRO something_to_clean;

    %IF &do_or_not_var %THEN
        %DO;

%* create a batch file to delete the orphan work directories;
%* create it in the work directory of this program so it will be autodeleted
at exit;

DATA _NULL_;
    LENGTH directory drive $ 256;
    SET orphandirs;
    FILE "&workdir.\cleanwork.bat";
    IF _N_ EQ 1 THEN
        DO;
            drive = COMPRESS(SCAN("&rootwork",1,':') || ':');
            PUT drive ;
            directory = 'cd ' || SCAN("&rootwork",2,':') ;
            PUT directory;
        END;
    PUT "rmdir /s /q " dirname;
RUN;

OPTIONS XSYNC NOXWAIT;

DATA _NULL_;
    LENGTH cmdline $127;
    cmdline = ''' || "&workdir.\cleanwork.bat" || ''';
    CALL SYSTEM(cmdline);
RUN;

    %END;
%MEND;

%something_to_clean;

```

CONTACT INFORMATION

Further inquiries are welcome to:

John M. Wildenthal
 JPMorgan Chase
 Phone: (614) 217-5715 john.m.wildenthal@jpmchase.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.