

Efficiencies, Tricks and FAQs of the SAS PIV-Grid

Mark Ezzo
VINCI SAS Administrator



October 12, 2017



Poll Question



- What is your primary usage of the SAS Grid?
- A ETL
- B Data Preparation (SQL & SAS)
- C Analysis (including Modeling)
- D Other



FAQs and Common Issues

* * * * *

- Data Efficiency Techniques
- Coding Techniques
- Parallel Processing
- Enterprise Miner
- Please ask questions at any time



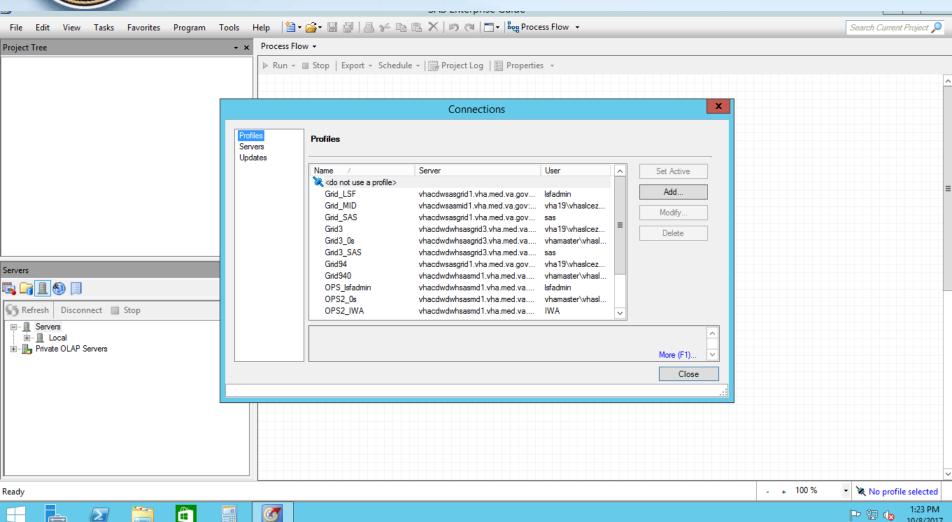
Grid Access/Connection Issues

When you do not see SASApp or any Servers, you are not connected. The reason is that you have not selected the desired profile as illustrated in the next slide. Please go through the VINCI Portal to access the Grid or connect to vhacdwapp15.vha.med.va.gov for Operations. We will use Research Grid references for our examples. As you can see in the next slide, it details "no profile selected". This means you are local and NOT connected to the Grid. The slide after shows a proper connection to the Research Grid and the necessary configuration of the profile (vhacdwdwhsasmd2.vha.med.va.gov for Research and vhacdwdwhsasmd1.vha.med.va.gov for Operations). If it says you are connected as a "Public User" you have not been configured in SAS Metadata. Please inform the VINCISASAdmins@va.gov and we will construct this for you.



No profile selected *

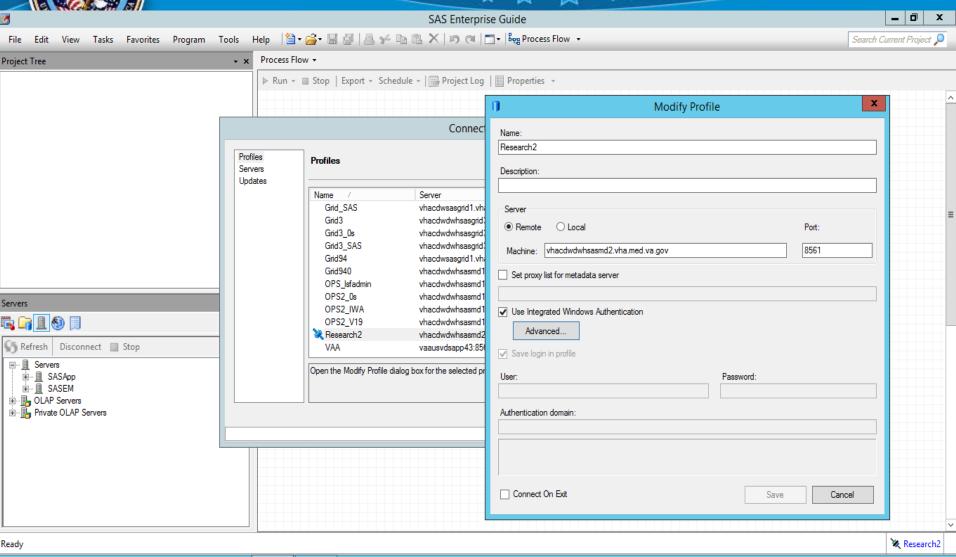




10/8/2017



Solution: Select the Profile



10/8/2017



Data Access and Efficiencies

- Avoid double clicking any table icon in the EG client which happens to point to a remote SQL Server data object.
 - Use SSMS application instead or
 - Run the Top 1000 Rows stored process in the EG application written by SAS admins.
 - You can used save librefs to load data to your Dflt schema in your project.
 It is a simple data lsqlserveribref.tablename;set saslibref.tablename;run;
- Avoid overusing the libref definitions which point to remote SQL Server databases.
 - Explicit SQL pass-thru is industry proven best practice for pulling records from RDBMS which don't have full in-database capabilities.
 - See posted "Best Practices in Accessing SQL Server Data" document on VINCI Central or SAS/Grid SharePoint site.
- Start each day with a fresh SASApp session when using the EG client application.
 - Current VA-wide policy forces the Kerberos security ticket to expire after 10 hours, Grid platform is hostage to this setting.
 - An expired Kerberos Tickets lead to the "unspecified GSS error".



Enterprise Guide Tips



- You have many options within EG to help facilitate usage.
- We will examine them under "Tools"
- Use the Programming Wizards under "Tasks" to learn and facilitate your programming
- LIVE Demo!



GSUB USES:



- Large Jobs
- Long Jobs (especially overnight)
- Production Quality Jobs
- Use with Linux access Products (Putty, MobaXterm and WinSCP)
- Produces Log and LST (output) Files
- Can run anything you do in EG (ODS reports, parallel jobs, etc.)



SAS Gsub Example 🖈 🖈



- Here is the command we will be using for batch processing on the Grid:
- sasgsub -GRIDSUBMITPGM /data/prod/admin/VHASLCEZZOM/Programs/Program.sas
- It breaks down as follows:
- sasgsub is the command
- -GRIDSUBMITPGM is the option to submit a program
- /data/prod/admin/VHASLCEZZOM/Programs/is the directory where the program is located
- Program.sas is the program you wish to run
- Live Demo!



Scheduling within the Grid

- You are able to schedule your jobs within the Grid utilizing a combination of crontab and a sas_cron script:
- #!/bin/bashcd /data/gridshare/\$USER.
 /app/shared/platform/Platform/lsf/conf/profile.lsf/app
 /shared/sasconfig/gsubconfig/Lev1/Applications/SA
 SGridManagerClientUtility/9.4/sasgsub gridwaitresults -gridappserver SASApp gridsubmitpgm
 /data/admin/VHASLCEZZOM/Programs/Test_gsub_s
 cript.sas
- Remember that all scripts in Linux are casesensitive!



How to access crontab





- Access putty_cac within your PIV account and edit via VI.
- [vhaslcezzom@vhacdwdwhsasgs2 ~]\$ crontab -l
- #* * * * * /data/gridshare/vhaslcezzom/Scripts/sas_cron; touch /data/gridshare/vhaslcezzom/sas_cron_has_run
- [vhaslcezzom@vhacdwdwhsasgs2 ~]\$ crontab -e
- crontab: installing new crontab
- [vhaslcezzom@vhacdwdwhsasgs2 ~]\$ crontab -l
- * * * * * /data/gridshare/vhaslcezzom/Scripts/sas_cron; touch /data/gridshare/vhaslcezzom/sas_cron_has_run
- [vhaslcezzom@vhacdwdwhsasgs2 ~]\$ crontab -e
- #* * * * * /data/gridshare/vhaslcezzom/Scripts/sas_cron; touch /data/gridshare/vhaslcezzom/sas_cron_has_run
- **Live Demo!**



Parallel Programming *



- This is one of the true strengths of any Grid.
- It allows for concurrent processing where you can run several jobs simultaneously, saving hours of processing time.
- You can use this for both Data Development and Analysis.
- Very simple to do with Grid Code Analyzer.
- Live Demonstration!



Enterprise Miner Uses



- Analysis
- Model Creation
- Model Testing (Interpolation)
- Model Publishing
- Can be shared with Project Group
- THE Leading-Edge Tool
- Automatically Grid-Enabled
- Live Demo





- SQL Pass-Through Queries
- SAS works faster and is more robust than SQL Server
- Ability to use SQL (Proc SQL) or SAS Data Steps and functions
- Work within the appropriate Venue (as stated previously)
- You can incorporate other technologies (i.e. R) within SAS (very desirable)
- SAS can create the data for other packages



More Tips



- Parallel processing can greatly reduce time in both ETL and Analysis.
- Far more storage in the Linux Grid than on Windows per project
- Standardized environment makes efforts simple and efficient
- Pre-prepared security and data via Metadata (always have your project enabled within Metadata)



Resources Offered 🙀 🖈 🖈

- 1 1 1
- The VINCI SAS Admins offer online training for you or your group.
- You can contact <u>VINCISASAdmins@va.gov</u> for SAS Specific questions or <u>VINCI@va.gov</u> for access, network, data or other questions.



Good Gridding!







Please contact VINCI SAS Administrators:

VINCI SAS Admins VINCISASAdmins@va.gov with any questions or comments.