Basic Research and Preparation (and a bit more) with the SAS Grid

Mark Ezzo

Data Scientist/

VINCI SAS Administrator

December 17, 2020

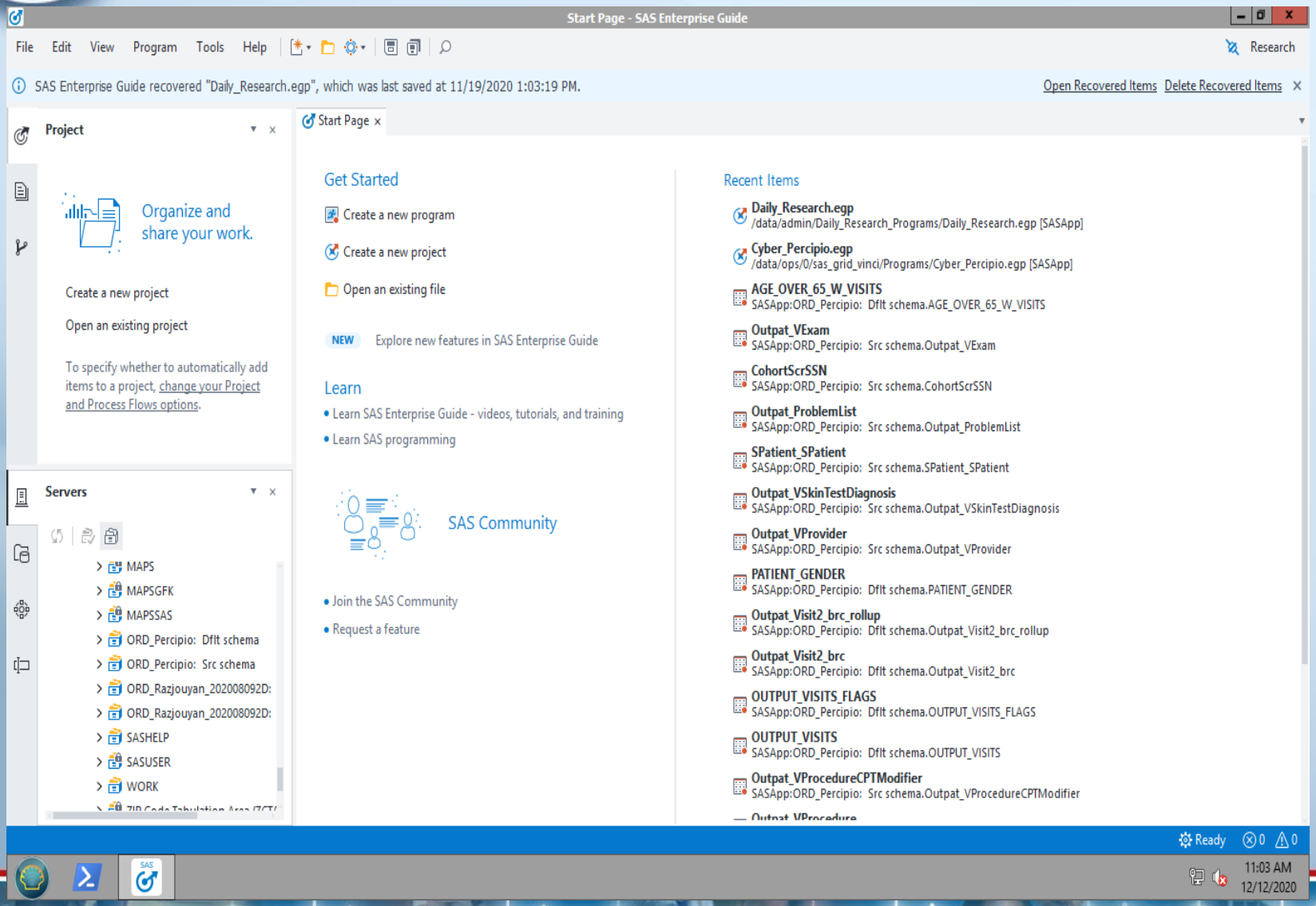
Seminar's Focus

- How to best approach a project via the SAS Grid
- Usage of Desktop SAS
- Limitations of ***NOT*** using Grid Resources
- Increased utilization of SAS EG
- Available code-generating Wizards
- ETL
- Basic Analysis
- Advanced Procedures
- Questions or Comments

Basic Assumptions

- All Paperwork is completed and approved
- Active Directory Groups are created and populated
- Data Managers have created the appropriated tables and views (.src and .dflt schemas) in SQL Server
- SAS Admins have provisioned the study in both metadata and Grid Storage
- You have a developed your hypothesis for acceptance or rejection (even implicitly)

Provisioned Study



Advantages of SAS Grid

- Desktop SAS is obsolete and has not been updated since SAS 9.X inception
- No updates nor enhancement will be forthcoming
- Not nearly or remotely the robust environment of the Grid
- Few programming aids
- Don’t use it

Advantages of SAS Grid

- Robust Environment of three compute nodes to enable parallel processing and accommodate a great number of jobs
- Terabytes of storage as opposed to only 300 GBs
- Greater memory allocation per job
- Enclosed Linux network (infinitely less interruptions allowing completion of jobs reliably)
- 3.7TB of workspace per compute node
- Additional security
- Award-winning UI (SAS EG and EM)
- Dedication SAS and Linux support staff

ETL

- Always use Proc SQL pass-through when accessing SQL Server tables unless they are very small.
- Wizards can create templates and teach.
- Customizations can be added to any generated code.
- Please consider space considerations and only save what is necessary!!!
- Demonstration

Analysis

- Many Wizards are available.
- You can customize them too.
- Create as a template for macros, etc.
- Demonstration

Graphs

- Graph Wizards are also available
- Can generate results for publication of code for use in programs
- SAS Graph via Wizards is simplicity itself and can be accomplished completely without coding
- No need to learn Graphics Language
- Demonstration

Our Current SAS Grid

Capability	Why it Matters
Centralized/Shared Workload Management	Effectively manage jobs and users, prioritization, compliance/auditing
High Availability	Avoid user or service disruption, rolling maintenance
Distributed Processing	Improved performance, meet changing demands, grow incrementally
Leverage Commodity Hardware	Reduce Costs

Live Demos

- Proc IML with R on the Grid
- Enterprise Miner with Open Source

- Architecture that is enabled by wide range of capabilities
  - End-to-end data cleansing/data mining and machine-learning process with a comprehensive, visual (and programming) interface

- Extension of the success of SAS GRID @ VINCI
  - Today the VINCI Grid utilizes SAS via user interfaces such as Base SAS, SAS Enterprise Guide, and batch submit

Copyright © SAS Institute Inc. All rights reserved.

9.4 &

- It's an enhances current strategy (Viya extends 9.4)
- Each is designed to solve different use cases
- Can co-exist
- Data, models and code can be accessed directly
- In 9.4M5, it's just another library engine

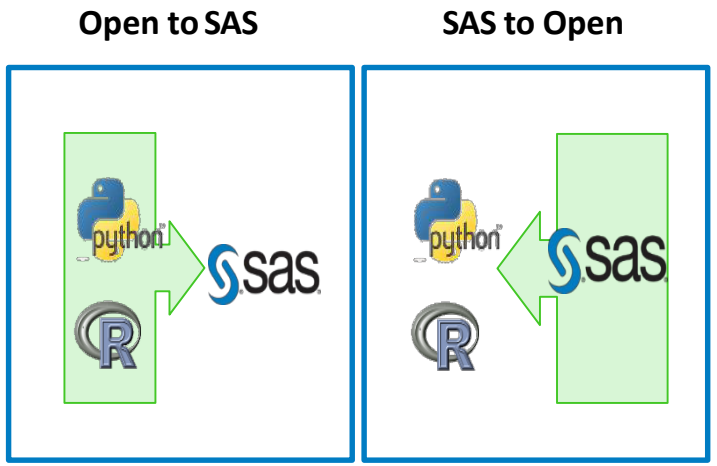
Copyright © SAS Institute Inc. All rights reserved.

SAS AS AN ENHANCEMENT

AND

SAS can augment open source

- Increase productivity
- Leverage your assets, people and platforms
- Bring the power of SAS to open source
- Create deployable analytics
- Goal is to ‘embrace’ and ‘extend’



Copyright © SAS Institute Inc. All rights reserved.

FOR MORE INFORMATION - EXTENDING R

Video: *Using R in SAS Enterprise Miner*
https://www.youtube.com/watch?v=TbXo0xQCqDw

Blogs: *Spectral Clustering in SAS® Enterprise Miner™ Using Open Source Integration Node*
https://communities.sas.com/docs/DOC-8011

Blogs: *How to execute a Python script in SAS® Enterprise Miner™*
https://communities.sas.com/docs/DOC-10832

Blogs: *Open Source Integration Using the Base SAS Java Object*
https://communities.sas.com/docs/DOC-10746

Article: *The Open Source Integration node installation cheat sheet*
https://communities.sas.com/docs/DOC-9988

Usage Notes:
http://support.sas.com/dsearch?Find=Search&ct=&qt=open+source&col=suppprd&nh=25&qp=&qc=suppsas&ws=1&qm=1&st=1&lk=1&rf=0&oq=&rq=0

Copyright © SAS Institute Inc. All rights reserved.

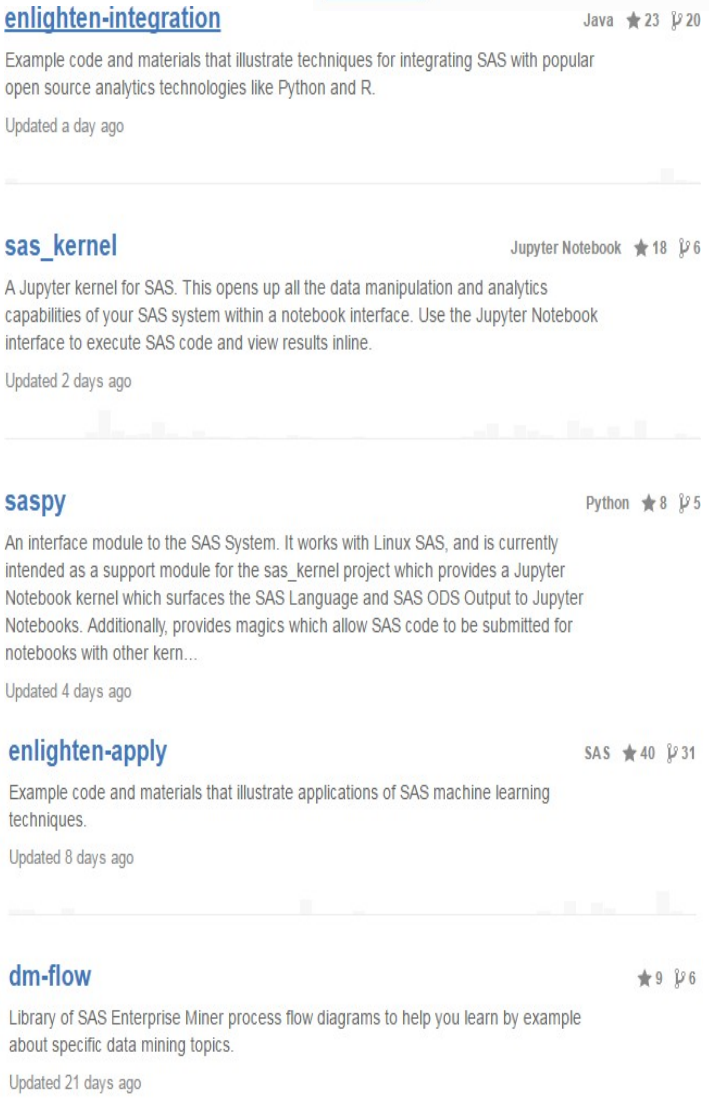
FOR MORE INFORMATION MATERIALS ON GITHUB

Sas integration and sample code
Integration with R, Python
https://github.com/sassoftware/enlighten-integration

Integration with Jupyter Notebook and Python
https://github.com/sassoftware/sas_kernel
https://github.com/sassoftware/saspy

Sample codes of SAS Machine Learning methods
https://github.com/sassoftware/enlighten-apply

SAS Enterprise Miner process flow diagrams
https://github.com/sassoftware/dm-flow



Copyright © SAS Institute Inc. All rights reserved.

Resources Offered

- The VINCI SAS Admins offer online training for you or your group.
- You can contact VINCISASAdmins@va.gov for SAS Specific questions or VINCI@va.gov for access, network, data or other questions.
- *VINCI Central site: http://vhacdwweb02.vha.med.va.gov/vincicentral/projectsites/SASGrid/Shared%20Documents/Forms/AllItems.aspx*

Good Gridding!

Thank you for attending.

Please contact VINCI SAS Administrators:

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