

## **CERES Systems Engineering Committee**

Members: Tammy Ayers, SSAI (DM)  
Denise Cooper, chair, SSAI (DM)  
Tonya Davenport, SSAI (ASDC)  
Vertley Hopson, SSAI (ASDC)  
Lisa Link, SSAI (ASDC)  
Walt Miller, SSAI (DM)  
Sue Sorlie, SSAI (ASDC)  
Scott Zentz, SSAI (SA)

Charter: Serve as a forum for resolving issues which affect more than one working group. Report to the CERES Data Management Team.

### **September 5, 2007 11:00 am**

The following members were present for this meeting: Tammy Ayers, Denise Cooper, Tonya Davenport, Walt Miller and Sue Sorlie. Other Data Management Team Members attending the meeting: Lisa Coleman.

The team reviewed the Shared Directories for Multiple Subsystems/PGEs issue.

Item 1: Shared directories for multiple Subsystems/PGEs. Jeff Walter had sent out an e-mail with the idealized directory structure for automation (see below). Many of the team had not had time to review his proposal. Since Jeff was unable to attend the meeting it was decided that the meeting would be held via e-mail with everyone sending out their comments/questions on Jeff's proposed directory structure. Everyone would respond with comments by Wed. Sept. 12, 2007.

The meeting was adjourned at 11:10 AM.

## **DRAFT**

### **ASDC Proposal for Changes to the CERES Runtime Environment**

#### **Introduction**

Over the many years that the CERES project has been operating, the manner in which data production is performed has evolved to be flexible and responsive to changing science requirements and priorities. However, the lack of a unified data production system to provide a foundational context to these efforts has resulted in a high degree of complexity in data management, testing and operations. As part of the ASDC Evolution effort, a system is under construction to automate CERES data production. Such a system requires some measure of standardization in the way certain things are organized and the way certain tasks are performed. The goal of this document is to propose some changes that could be made to reduce the complexity of the current paradigm and make

the automated system easier to implement without placing an undue burden on the CERES SCF or ASDC staff.

### **Data Organization**

Currently, data locations on the production system staging disk areas are organized by the subsystem that produces them. If one subsystem requires the output of another subsystem as input, then that subsystem must have a priori knowledge of the other subsystem's directory structure. In some cases, there may also be ancillary files that are shared by multiple subsystems. This high degree of coupling could potentially lead to problems if one subsystem decides to reorganize its data directories. It also makes management and tracking unnecessarily burdensome.

To eliminate this coupling and facilitate the implementation of the automated system, ASDC proposes completely separating data from the subsystems that use and produce them and flattening out the existing structure. The first step in this process is to reorganize all data under a single data root directory. Within this data root directory, data would be organized by "data type" rather than by subsystem. As ASDC works on integrating various PGEs into the automated system, data type names are being generated that should hopefully prove to be self explanatory and intuitive. For the most part, the CERES data type names are created from the first part of the CERES data product file names. Other data type names are created to reflect what they are but they may not exactly match the first part of the file name. Organizing the data in this way makes it much easier for the automated system to manage the staging disk area and to control the flow of data into and out of the system.

Also contained under the data root directory would be a separate directory for organizing static files. The intent of this directory is to make the static files permanently resident on the staging disk with backup copies stored in the archive. The organization of the static directory is optional. ASDC recommends organizing the files by "data type" in the same manner as the dynamic data. However, if it makes more sense to organize them by subsystem or PGE, this can be easily accommodated as well. The intention would be to place all MCF files in a single directory under the static directory. But, again, this is optional and the MCFs can be organized by subsystem if that makes more sense for some reason.

An example of the proposed structure is provided below. Obviously not all data types are shown and the location of the data root directory is simply an example.

```
/gpfs/data/root/  
    CER_HQCP/  
    CER_CRSVB/  
    CER_HMAVAIL/  
    CER_HMSAL/  
    CER_MOA/  
    CER_PQCR/  
    D4FAXMIS/
```

D4/FAPMIS/  
SNOWf13NORTH/  
SNOWf13SOUTH/  
PH/ # For production history tar files

Static/  
Ozwts/  
GridParams/  
MCF/

### **Executables**

Since nearly all CERES PGEs are currently run manually at the ASDC, it is not surprising that most of the PGE scripts provided to ASDC by the SCF appear to be geared toward this model. However, the automated system will perform most of the tasks that are currently performed by the existing scripts including PCF generation, handling of log files, and cleanup of temporary files. To facilitate automation and reduce the complexity of testing and operations, the existing scripts should be stripped down and/or eliminated to create a uniform PGE interface with the automated system. Details are provided below.

- 1) PCF generation scripts – With the automated system, the SCF no longer needs to provide any PCF generation scripts or other related functionality. The automated processing system handles all PCF generation functions.
- 2) Top Level Executable Scripts – These scripts should still be provided to ASDC with PGE deliveries. However, much of their current functionality could and should be eliminated. Below is a list of things that the scripts should do, some things they should not do, and some things that are optional.
  - a. The scripts SHOULD
    - i. Be responsible for calling the appropriate binary executables. However, as shown in Section 3 below, it would be preferable if any hard coded or semi-hard coded paths to the binaries be removed.
    - ii. Generate an appropriate exit code for the PGE and exit the script with that exit code.
  - b. The scripts SHOULD NOT
    - i. Set the PGS\_PC\_INFO\_FILE environment variable. This is done by the automated system using the PCF name that the system generated.
    - ii. Execute any script to perform any kind of PCF generation or manipulation. Again, the automated system will handle all PCF related matters.
    - iii. Do anything with log files or the PCF upon PGE completion. The automated system tars up all runtime logs, archives them, and places them in a directory in the data staging area if access to them is needed in the near term following the PGE run.
  - c. The scripts MAY (if desired)

- i. Send an email notification to the CERES SCF upon completion. However, ASDC recommends removing all hard coded or semi-hard coded paths in the script. The automated system creates a temporary directory for each PGE execution that serves as the current working directory for that job. Any non-product files that need to be written out can simply be written to “./”. It is also not necessary write a file to disk containing the contents of the email message. A file handle can be created that is a direct pipe to your favorite mail command.
- 3) Binary Executables – The binary executables should be portable. In other words, they should be able to run from anywhere and not be tied to a particular directory structure. In the automated system, the preferred way to organize the PGEs is to have all PGE executables (both scripts and binaries) and their associated S4PM configurations organized by PGE name and SCCR number under a central PGE root directory. An example is given below using the MOA PGE.

/gpfs/pge/

CerPGE121P1/

SCCR615/

CERPGE121P1.ancillary.cfg

CerPGE121P1.pcf.tpl

CerPGE121P1\_PROF1.cfg

MOA\_Gen.exe

runmoa.pl

The files MOA\_Gen.exe and runmoa.pl are the PGE binary and top level run script, respectively. The remaining files are the S4PM configuration files and PCF template file. Organizing the PGEs in this structured “drill-down” type of way serves the following purposes.

- a. It provides a uniform structure that facilitates ASDC’s ability to manage and test them.
- b. It provides the automated system a method for selecting the appropriate PGE on the fly using a single body of code. By specifying the PGE name and SCCR number in the production request form, the automated system is able to construct the path to the appropriate PGE name, version, and configuration that are required.
- c. It provides a fallback mechanism if this is ever required. Previous versions of a particular PGE can be left out on disk as long as desired without interfering with the execution of newer versions. The selection of the appropriate version to run is provided in the production request.
- d. Preservation of older PGE executables and binaries could be useful for troubleshooting problems. If an issue is discovered with a particular science product at some later date, preservation of this information provides additional traceability.
- e. It facilitates configuration management. All PGE configurations and executables will be preserved in Subversion.

### **Coordination with the SCF**

As the ASDC works on integrating the CERES PGEs into the automated system, regular communication and coordination with the CERES SCF will be paramount. Numerous questions are already being generated in the early stages of the integration process.

ASDC requests the following items from the SCF.

- 1) Point of Contact (POC) information for each of the CERES subsystems – As questions come up, it would be very helpful to have a specific person to contact. Depending on the complexity or weight of the problem, some issues and questions may be able to be handled via email or phone call and some may require a meeting. This will have to be decided on a case by case basis.
- 2) At least one meeting with each of the subsystem leads – The purpose of this meeting would be to review the implementation, and to make sure all questions have been answered and all issues have been addressed. It is expected that the SCF will have questions regarding modifications to the top level run scripts. It is a given that the ASDC will have questions and will need the SCF to verify that certain things are being done correctly by the automated system.