U.S. Department of Energy Best Practices Workshop on File Systems & Archives

San Francisco, CA

September 26-27, 2011

Position Paper

NCAR's Centralized Data Services Design

Pamela Gillman
NCAR Computational & Information Systems Lab
pjg@ucar.edu

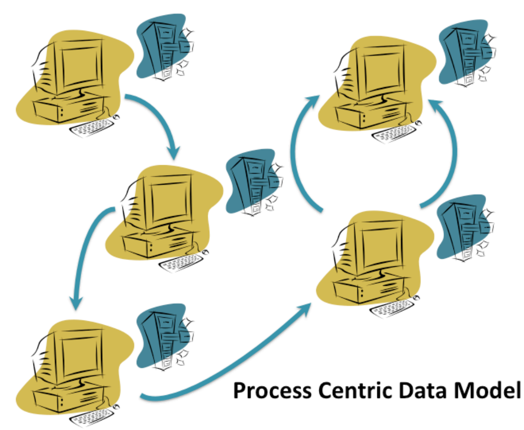
Erich Thanhardt
NCAR Computational & Information Systems Lab
erich@ucar.edu

ABSTRACT / SUMMARY

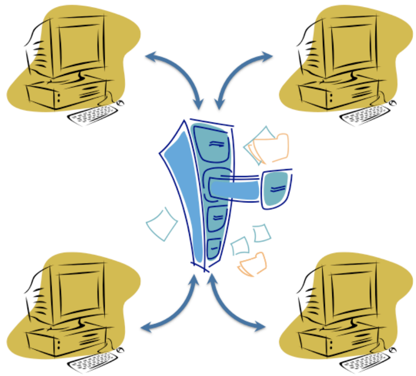
As NCAR designs and builds our next generation computational center, we are exploring ways to evolve scientific data workflows from a process centric model to a more information centric paradigm. By looking at cyberinfrastructure design, resource allocation policies and software methodologies, we can help accelerate scientific discoveries possible from computational resources of this scale. We will explore the challenges we have identified in our data architecture and present some of our current projects moving us towards an information centric solution.

INTRODUCTION

Computational centers have traditionally provided systems architected for a single use such as computation, data analysis or visualization. Each resource has local storage configured for the typical task performed and the center provides a common tape based archival system. This encourages a scientific workflow that typically involves retrieving input data from the archive system, generating new data, including intermediate files necessary for the next run, and finally storing all data back on the archive system. If you wish to post-process, analyze or visualize the data, you need to read the data back onto a different resource. This process may repeat multiple times as either issues are identified in the data, or discoveries spark a new direction of research.



When surveyed, users identified the movement of data between resources as a significant bottleneck in their workflow. Therefore, armed with this information, and driven by the ever-escalating costs of archival systems and the increases in ability to produce data, we started looking at architectural solutions to evolve workflow. The traditional workflow model is very process centric. Data moves between resources dedicated to a single step in the overall process and the archive essentially becomes a file server. What if we started looking at the data as the center of the workflow? Not only would this decrease the number of data movements in the workflow, but it can potential decrease the amount of data ultimately targeted for actual archiving. We now refer to this as an information centric model.



Evolving Scientific Data Workflow

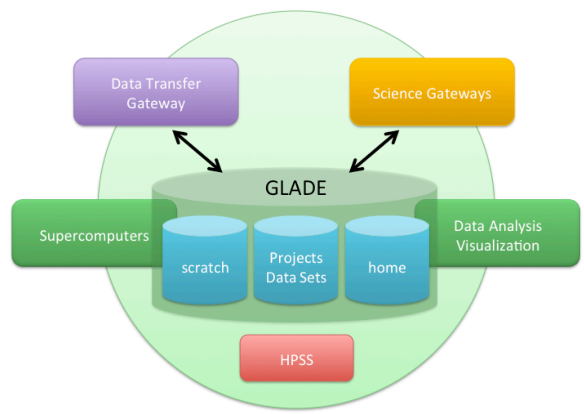
Based on an analysis of current workflows, we identified several challenges in evolving data workflow toward the new paradigm. Many bottlenecks exist in current workflows; it's time consuming to move data between systems; bandwidth to the archive system is insufficient; and available disk storage space is insufficient.

This presents a bit of a 'chicken and egg' problem. The current environment potentially shapes current user behavior. How do we anticipate the behavioral changes that will occur with a significant change in the environment? Storage cost curves are steeper than compute costs so how do we find the right balance between storage and compute investments? Archive costs are on an unsustainable growth curve so how do you better balance usage of disk space versus archive space?

Globally Accessible Data Environment

The first step we took was to centralize our data analysis and visualization resources around a centralized storage system. This was a lower bandwidth parallel file system solution that provided users of these resources a single namespace for their work. Space was provided not only for short-term 'scratch' usage but also for longer-term project spaces necessary for data analysis and visualization work.

The next step was designing a scalable architecture that encompassed all HPC resources including access to data collections managed by NCAR. This architecture needed to be flexible enough to support current systems and science gateways, and also able to scale as HPC resources grow with the new data center. And to ensure the shift in workflows, the user needs to be able to interface with this environment in the same way no matter which resource they are working from or what task they are trying to accomplish. High bandwidth connectivity to any resource within this environment and a choice of interfaces support current projects and are flexible enough to support future requirements.



The GLADE data architecture becomes the centerpiece of the new information centric workflow. Data can stay in place through the entire process as GLADE provides not just 'scratch' space for computational runs, but persistent longer-term storage for data processing, analysis and visualization. This persistent storage allows completion of the entire workflow prior to final storage of results either at NCAR or offsite. The addition of high-bandwidth data transfer services with direct access to the GLADE environment provides efficient methods for moving data between NCAR and peer institutions.

Accounting Systems Enhancements

To allow for better tracking of resource usage, the NCAR accounting system is being redesigned to account for computational resources, data analysis and visualization resources, disk storage usage, data transfer services and archival services. These tools will allow NCAR to fine tune the balance between resources based on evolving usage patterns due to changes in workflow.

Workflow Examples

Case 1: Nested Regional Climate Model (NRCM)

The project group has common access to 'scratch' space and a dedicated longer-term project space. The computational team submits a model run to the supercomputing queue. The model outputs approximately 100 variables per time step along with intermediate data files associated with startup of the next time step to the 'scratch' file system. Once the model completes, a post-processing job pulls approximately 20 variables of interest into data analysis files writing these into their project space. The analysis files are now available for analysis by the research team. This data will stay in place as long as necessary to complete the analysis. A final job step writes the final full output files to the HPSS archive. Since the 'scratch' file system is purged regularly, intermediate files that are no longer needed are never stored on the archive, and smaller data sets needed for analysis are available to the team right after the computational run completes.

Case 2: Research Data Archive (RDA)

The Research Data Archive provides access to common data sets to the research community. Prior to the implementation of the GLADE architecture this data was only available from the archive system. By allocating space within GLADE for the RDA data, access can now be granted directly from NCAR's HPC resources. Previous workflows needed to first copy this data from the archive to a 'scratch' area before running the computational job. There were costs in time required to access the data, space required to hold the data and the side effect of numerous copies of the same data being on disk at the same time. With direct access now available, jobs use the data from a central location that's immediately available to all jobs and doesn't rely on archival access..

CONCLUSIONS

We feel that we have made progress towards a better architecture to meet the diverse needs of our user community. We believe that this architecture is sustainable into the future and will help balance the costs associated with compute/storage/archival. Checks are in place so adjustments can be made as user behaviors change and hopefully data management becomes not just a tedious task, but also something that results in more productive scientific discovery.