TRI Data Storm

William P. Baird, Wes Bethel, Jonathan Carter, Cristina Siegerist, Tavia Stone, and Michael Wehner

NERSC/CRD, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA, 94720

Introduction

TRI Data Storm centers around the concept of using an integrated, multisystem file system to improve the analysis of results produced by the demanding HPC application - the community atmospheric model (CAM). Our aim is to take the output of CAM from a high-resolution grid, filter out the data of interest and visualize the formation of storms in the North Atlantic basin. This tool will be used in a study comparing real hurricane data with simulations. The effect of climate change on the intensity and frequency of hurricanes in area is of utmost importance to policymakers.

While this is a fairly generic workflow that could hold true for virtually any HPC application, the unique aspect to our approach is that there is a single high-performance parallel file system serving all of the different systems and applications. For TRI Data Storm we use an externalized WAN-GPFS (Wide Area Network General Parallel File System, a product of IBM), shared out by a dedicated cluster and mounted on all of the different computational resources used by our tool: an IBM Power system, a Linux cluster, and a SGI Altix. GPFS removes the necessity of transferring data between the different systems manually, and allows the most appropriate hardware for the task at hand to be used more easily and effectively. It also transparently addresses the security and authentication issues of using remote systems. In addition, as compared to other remote files systems, GPFS provides excellent performance and scalability. It presents no bottlenecks to the individual applications, and effectively links our compute platforms together.

Methods

The CAM output is spread over many netCDF files that contain considerable data that is not important for the visualizations considered here, so we must filter out the data of interest and then pass it on to the visualization application. Our whole analytics process is composed of three components. The first two, CDAT and AVS/Express, are both conventional application software. The final, GPFS is the software that enables us to move data to CDAT, and from CDAT to AVS/Express, efficiently and seamlessly.

CDAT has the capability to read a netCDF dataset and write out only a specified subset of the fields. By launching multiple copies of CDAT that can operate independently on different netCDF files as an embarrassingly-parallel application on a Linux cluster we can rapidly process the CAM output to produce just the data needed for the visualization step.

As a starting point in our visualization application we used the AVS/Express built-in netCDF reader to read the datasets output by CDAT. Using a GIS world map we superimpose the dataset to the surface of the world coloring by the pressure at surface level. We also show wind velocity vectors as arrows, and an elevated surface of the precipitable water in the region close to the hurricane eye. The datasets are read in using a loop through the time steps and a sequence of images output to later assemble into an mpeg movie. So far we have used built-in visualization capabilities, but using the module development API we will add custom analysis modules to the data flow.

GPFS is a cluster file system like many others. The difference is that through the multicluster capabilities that IBM has introduced, the file system can be exported to multiple independent systems. Additionally, GPFS has the ability to share the same file system between different operating systems: both AIX and Linux are currently supported. In this case, an owning cluster is set up independent to all the other computational assets. The file system is then imported to a massively parallel computer (an IBM Power system) at NERSC, a commodity Linux cluster at NERSC (Intel nodes with Ethernet interconnect), and a CC-NUMA SMP with graphics hardware (an SGI Altix) located on the floor of SC ’05. Each step of the computational process reads the required data set from the file system. The method of data transport depends on each of the systems. The SP and Linux cluster are connected via Ethernet and the data is transferred via NSD servers. The SGI Altix takes advantage of a feature in GPFS that allows clients to be connected via fibre channel to the storage giving much faster access to the data than over IP-Ethernet connectivity. Ethernet connectivity is still required to the servers, but the bandwidth through this 'local discovery' mode is far greater. With these capabilities we are able to connect together all of these disparate systems at high bandwidth.

Figure 1 Connectivity diagram

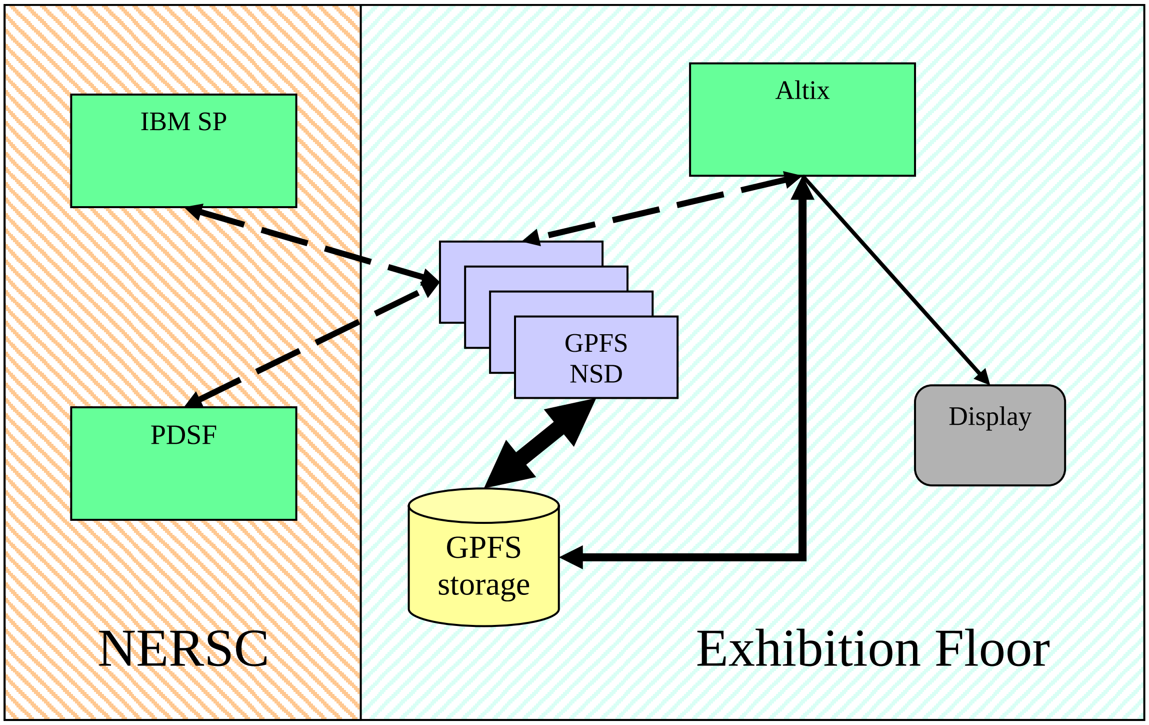

Figure 1 shows how the systems are connected, and how they are distributed. Systems in the orange striped box are located at the NERSC center, while systems in the blue striped area are located on the exhibition floor of SC ’05. Solid connections indicate fibre-channel links, while dashed connections are Ethernet.

Figure 2: Workflow diagram

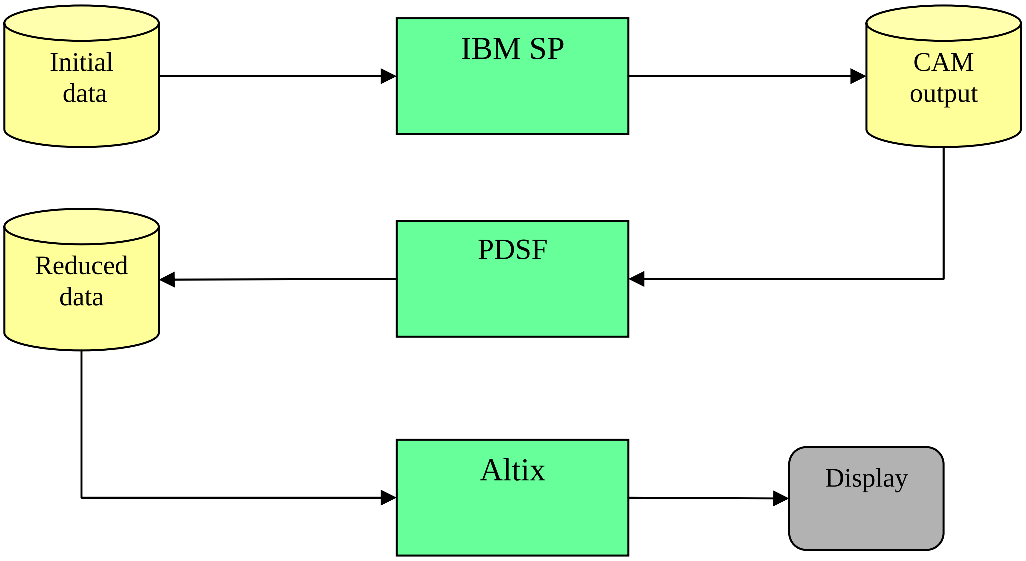

Figure 2 illustrates the workflow for this project, showing the central role played by WAN GPFS. A data set of initial conditions, such as ocean temperatures, is used as input to CAM running on the NERSC IBM Power resources. During the course of the simulation, a large number of history files containing atmospheric grid-point data are. The Datasets section, below, describes the resolution and format of these files. The output data is then post processed to produce intermediate files containing only the data to be used in subsequent steps. This data reduction step is run on a commodity Linux cluster at NERSC using the CDAT toolkit to extract the precipitation, surface pressure, and wind vector. The final step of the process is the analysis and visualization of the data. For TRI Data Storm, we use CDAT, a toolkit written especially for manipulating and visualizing climate data, and AVS Express, both running on an SGI Altix for the visualization and analysis. In this case, the visualization is of the precipitation, pressures, and wind vectors of a hurricane.

Applications and Tools

The Community Atmosphere Model (CAM) is the atmospheric component of the flagship Community Climate System Model (CCSM3.0). Developed at the National Center for Atmospheric Research (NCAR), the CCSM3.0 is used to study climate change. The CAM application is an atmospheric general circulation model (AGCM) and can be run either coupled within CCSM3.0 or in a stand-alone mode driven by prescribed ocean temperatures and sea ice coverages. AGCMs are key tools for weather prediction and climate research. They also require large computing resources: even the largest current supercomputers cannot keep pace with the desired increases in the resolution of these models. CAM 3.0 produces a series of NetCDF-format history files containing atmospheric gridpoint data generated during the course of a run. History files contain model data values written at specified times during a run. One can specify the frequency at which the data is written. Options are also available to record averaged, instantaneous, maximum, or minimum values on a field-by-field basis. There are nearly 300 different model data values that can be written out.

Climate Data Analysis Tools (CDAT) is a software infrastructure that uses the Python object-oriented scripting language to link together separate software subsystems and packages, forming an integrated environment for: management of gridded data, large-array numerical operations, and visualization. CDAT was developed by the Program for Climate Model Diagnosis and Intercomparison at Lawrence Livermore National Laboratory.

AVS/Express is an interactive framework that enables rapid prototyping of complex 2D and 3D visualizations. It is a modular, hierarchical, open and extensible system that offers an extensive library of modules to build visualization applications. It includes visualization modules, user interface creation modules, interactivity support modules, application development modules and I/O modules. It also provides a C, C++ and FORTRAN API to develop custom libraries.

Datasets

Our CAM simulations use a high-resolution grid, defined as 1000x721x27 gridpoints (longitude, latitude, and elevation). There are 292 different model quantities written out every hour of simulation time, and about half of these quantities have no dependence on elevation. Each quantity is written out as a separate netCDF field. In total, there are 3530 32-bit reals written for each longitude and latitude point, requiring 9.5 GB per hour simulated. We estimate that we will analyze a simulation of 10 days, giving a total dataset size of over 2 TB. The netCDF datafiles are each less than or equal to 2 GB in size, each snapshot in time spanning multiple files.

In the post-processing step we will condense the data and retain only the fields to be visualized: surface pressure, wind speed, and precipitable water.

Acknowledgements

This work was supported by the Director, Office of Science, Office of Basic Energy Sciences, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. This research used resources of the National Energy Research Scientific Computing Center, which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.