CLIMATE DATA PROCESSING MADE CHEAP

Scott Mindock

Space Science and Engineering Center
University of Wisconsin-Madison
Scott.Mindock@ssec.wisc.edu

Climate data processing can be done cost effectively. Cost savings can be achieved by leveraging commodity hardware and open source software. A system can be integrated by applying standard web technologies. The Atmosphere PEATE design team at the University of Wisconsin has developed a system following these strategies. The Atmosphere PEATE employs open source software packages and commodity hardware. Standard Web Service technologies have been applied to realize the PEATE as a Service Orientated Architecture. The system was built cost effectively in both hardware costs and software development time.

The NPP Atmosphere Product Evaluation and Test Element (PEATE) is composed of commodity hardware. The system utilizes a Linux cluster and several file servers. Clusters have been proven to be a reliable way to tackle large processing jobs. The modularity of a cluster allows the system to be upgraded and expanded over time. Processors and memory prices tend to decrease over time, processing nodes that are cheaper and more powerful can be added as the system matures. The storage system is designed to utilize standard storage servers. Servers can be added or removed from the system without requiring the system to be stopped.

The Atmosphere PEATE is written in JAVA. JAVA was selected because it is a mature technology and has a large development community. The maturity of JAVA provides cost savings in many ways. JAVA development tools are readily available, cheap and reliable. JAVA hosts a large variety of third party packages that provide run time services such as logging, network communication and database connectivity. These tools and software packages provide excellent functionality at little or no cost. JAVA also hosts built in support for code generation, which is used to generate large amounts of system software, again, saving money.

 The PEATE team used several open source tools for software development. The Eclipse IDE was the backbone for software development. Eclipse provides a robust environment for JAVA development with literally thousands of plug-ins available to help developers. Apache ANT is a "Make"-like facility for JAVA simplifying the build process; coupled with JUNIT our team was able to create a suite of automated tests executed during our nightly build.

The PEATE system contains large amounts of generated code. XML based system configuration files are both read and written using code generated from the XML schema definition for the file.

Communications between subsystems contains generated interface code, hosted by Open Source transport layers.

Web Services are widely accepted and deployed in the software community. The PEATE is organized as a set of cooperating Web Services. This allowed the development team to eliminate the need to develop a custom set of communication protocols. Networking protocols are generally costly to implement and maintain; by leveraging Web Services technology, the project realized a significant savings in develop time.

Climate Data Processing requires the handling of very large datasets. PostgreSQL is an open source database requiring no licensing fees. The database provides a way to manage the large datasets. JDBC provides JAVA interfaces to access these databases using SQL.

The Atmosphere PEATE achieves a cost effective solution to the challenge of developing a Climate Data Processing system. The PEATE development team has done this by utilizing readily available hardware and Open Source Software packages   In other words Climate Data Processing has been made CHEAP.