Maximizing the Usefulness of Grid Technology on NOAA Office PC's

FY 2004 Proposal to the NOAA HPCC Program

August 11, 2003

| Title Page | Proposed Project | Budget Page |

Principal Investigator: __**Mike Kane**_____

Line Organization: NESDIS
Routing Code: E/SP1__
Address:

                NOAA/NESDIS/OSDPD/IPD
                RM 3065 FB4
                5200 Auth Road
                Suitland MD 20746

Phone: (301) 457-5165x119
Fax: (301) 457-5199
E-mail Address: Mike.Kane@noaa.gov

Other Investigator 1	Other Investigator 2	Other Investigator 3
Ingrid.Guch@noaa.gov	Jones@cira.colostate.edu	Carl.Karlburg@noaa.gov
Other Investigator 4	Other Investigator 5	
Ralph.R.Ferraro@noaa.gov	Kidder@cira.colostate.edu	

Proposal Theme: **Technologies for Collaboration, Visualization or Analysis- Enabling Applications**

Signature 1 *(required)*	*Signature 2* *(required)*	*Signature 3* *(optional)*
_____	Barbara A. Banks	Authorizing Official 2
Information Technology Specialist	Chief	Title
NOAA/NESDIS/OSDPD/IPD	Information Processing Division	Organization

Using Grid Technology to Harness Office PCs for Improved Satellite Data Processing and Technology Transition

Proposal for FY 2004 HPCC Funding

Prepared by: __Ingrid Guch (for Mike Kane)_____

Executive Summary:

Office PC's sit unused during evenings, weekends, meetings and lunch hours, which can easily exceed 80% of each week. A Grid using existing office PCs has the potential to be an enormous computing resource for NOAA, particularly considering current and planned connections of NOAA to the Next Generation Internet. In the FY 2002 HPCC funded project, "Harnessing NOAA Office PCs for Improved Satellite Data Processing and Technology Transition," a cluster of seven office PCs was set up and the proof of concept successfully demonstrated the ability to securely process satellite data during idle time. The purpose of this proposal is to demonstrate the expanded implementation of the concept in a manner that the system can be sustained reliably and at low-cost into the future. Specifically, the system will be placed on office PCs as they are upgraded at the NOAA/NESDIS/OSDPD Suitland, MD campus. Approximately one-third of the 120 office PCs are upgraded each year. A preliminary cost-benefit study revealed that if this were successfully done and the CPU cycles were used by applications more than $1,000,000 in hardware and associated staffing costs could be avoided over five years.

Work related to this will be funded by NOAA/NESDIS/OSDPD and NOAA/NESDIS/OSD to provide training for applications programmers as well as to provide a more complete cost-benefit analysis to study savings that may occur during the transfer of research to operations for PC-based software that comes from cooperative institutes and universities. This project has future technology transfer opportunities within NOAA Satellite and Information Services (NCDC, ORA), as well as within NOAA's National Weather Service (NWS) and NOAA's Office of Oceanic and Atmospheric Research (OAR). NOAA's ability to meet current and future obligations to customers in a timely and cost-efficient manner will increase.

Problem Statement:

(1) NOAA is continually seeking low-cost sources of additional computing cycles. For example NOAA Satellite and Information Services requires frequent increases in computing cycles to process satellite data streams in response to the variety of sensors, algorithms, formats, timeliness, and refresh requirements from operational users, as well as for reprocessing and parallel processing efforts to ensure climate-quality data. Computer systems running Polar Operational Environmental Satellite (POES) applications are frequently at or near100% capacity when the orbits from the operational afternoon satellite arrive, typically 14-15 times/day (figure 1). Each time the system is near 100% capacity processing slows down and the delivery of products and services often are at risk of not meeting timeliness requirements.
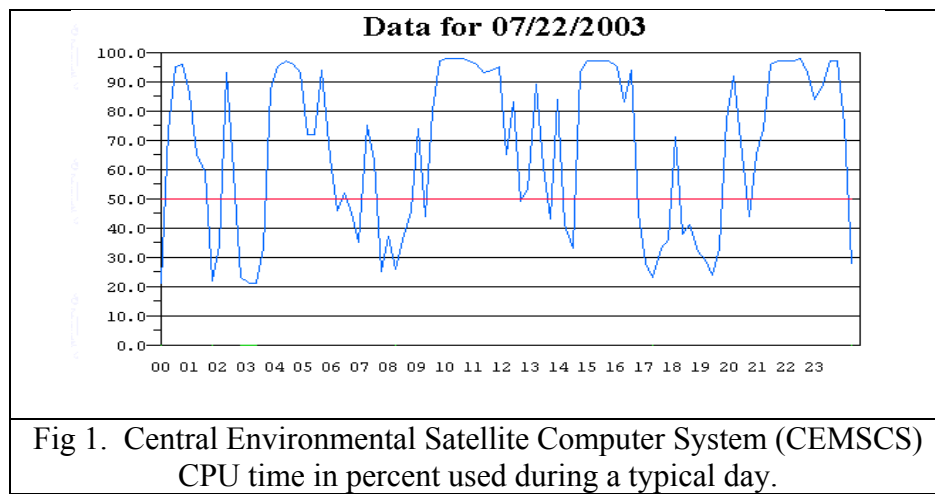

Fig 1. Central Environmental Satellite Computer System (CEMSCS)
CPU time in percent used during a typical day.

(2) The current domain for the cluster of seven office PCs, successfully demonstrated to be capable of processing POES applications, is maintained separately from other office PCs. The cluster domain must replace or be linked to the regular office PC domain to ensure that maintenance and upgrades are performed on the entire grid of all PCs, regardless of whether or not they are part of the cluster. Until this can occur, significant satellite data processing efforts will not be placed on the cluster for fear that the grid will not be properly maintained.

Transitioning the current seven PC cluster to a grid that can be cost-effectively sustained into the future and be accessed for processing satellite data in real-time or for reprocessing or parallel efforts clearly fits into the HPCC goal "… to enable information research and technology development that enables NOAA to provide greater access to its vast holdings…" The successful demonstration of this project will provide NOAA improved access to real-time information, complete and more accurate information for NOAA's historical archive, and capitalize on unused (idle) computing power at little to no additional cost to operations.

Proposed Solution:

The proposed solution is to transition the current seven PC cluster to a grid with up to 40 Office PCs that can be cost-effectively sustained into the future. It will be accessible for processing satellite data in real-time, for reprocessing or for parallel processing efforts. This grid will be running in NOAA/NESDIS/OSDPD.

Implementation Plan

(1) NOAA/NESDIS/OSDPD personnel in charge of maintaining the Office PCs will be trained to use a Windows 2000 Domain controller that is currently being used for the PC cluster. They will compare and contrast the capabilities with the current Samba-based domain controller being used for Office PCs.
(2) NOAA/NESDIS/OSDPD personnel in charge of maintaining Office PCs will generate a plan for the best domain structure to use. There are at least 2 options… (1) the cluster domain and the Samba domain could remain coexisting as they are now but have automated tools to ensure they are both upgraded at the same time or (2) the cluster domain could replace the Samba domain
(3) Contractors will implement the best domain structure outlined in the plan
(4) Contractors will develop a new image to use for initializing new office PCs that includes the Grid clustering set up, so that all new office PCs in NOAA/NESDIS/OSDPD can be accessed during normally idle time for satellite data processing.
(5) {From Matching Funds} NOAA/NESDIS/OSDPD, ORA, and NCDC personnel will be trained on how to incorporate applications into the Grid computing system.
(6) {From Matching Funds} CIRA and NOAA/NESDIS/OSDPD will conduct a cost-benefit analysis assessing the technology transfer improvements anticipated for PC-based applications moving from university or other environments to an operational status.

Analysis:

The system currently running on the seven PC Grid cluster is using the Data Processing and Error Analysis System (DPEAS) developed by CIRA with support from DoD and NOAA. DPEAS ingests data and distributes ("Parallelizes") the processing of them over a network of computers running Microsoft Windows 2000 and/or Windows XP. As an added benefit, DPEAS was built with optimized technology and research transfer in mind by being built to use the most common platform available, the office PC running Windows. DPEAS is stable and has been tested in a 20-processor configuration. Simulation tests have shown that it is capable of scaling to approximately 2,000 processors. DPEAS was designed specifically for processing satellite data from multiple sources and for both research and operational needs. DPEAS has hooks in place to take advantage of the Open Grid Services Architecture (OGSA) currently being defined based on the Open Grid Services Infrastructure (OGSI). Once secure FTP capabilities are standardized via these or other grid computing forums the DPEAS system could go across institutional or geographic boundaries with minimal modifications.

DPEAS can be easily integrated with other related packages for satellite data processing because it leverages NASA HDF-EOS format and technology for its main I/O manipulations.

One alternative to this approach to obtaining additional CPU cycles is to purchase a Linux-cluster of machines dedicated to satellite processing. It would take a cluster of 44 dual-processor Linux machines to have the same CPU power as currently available 100 Office PCs. Staffing and maintenance of this kind of Linux cluster would likely exceed $200,000 per year making it a less desirable approach.

Another alternative is a private PC-grid company. Currently, Entropia appears to be the leading company offering this service. Their product, DCGrid 5.0, starts at $50,000 plus an annual maintenance fee of $10,000 per year. However, their software does not support the satellite data processing capabilities or research to operations streamlining inherent in the DPEAS system, making it a less desirable approach.

The benefits of the proposed solution are that (1) most hardware and staffing costs can be avoided when using existing and maintained NOAA Office PCs for the additional CPU capacity and (2) the DPEAS software is already customized for NOAA, DoD, and NASA satellite data processing needs.

Performance Measures:

PM#1: The number of office PCs with the capability to process satellite data in their idle time will increase to a grid of 40 Office PCs in the first year (to 120 in 3 years).

PM#2: {related to matching funds}: The number of processing jobs running each day on Office PCs will increase by at least 10% in the HPCC demonstration year. The rate of increase each year thereafter will be a function of the number of PCs added to the grid, but should be at least 10% each year once the system is transitioned to operational status.

Milestones

Month 01	Funds in Spending Plan
Month 02	Personnel trained in Windows 2000 domains
Month 03	Plan for Office PC domain maintenance written
Month 04	{related to matching funds} Application programmers trained to access Grid
Month 05	Office PC domain plan implemented
Month 06	{related to matching funds} New applications running on the Grid
Month 07	System initialization image for new office PCs contains Grid capability
Month 08	{related to matching funds} Cost-benefit report assessing the transfer of research to operations using this system vs. traditional methods

Deliverables

- o Domain Maintenance Solution Plan
- o 40 Grid Office PCs on sustainable domain
- o PC initializing Image with Grid capabilities
- o Midterm report
- o Final report