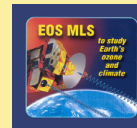
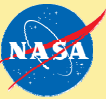


MLSgrid: Development of a System for more Efficient Science Data Processing

Brian W. Knosp¹, Paul Zimdars², Christopher W. Chambers³, and Susan Neely²



1 Jet Propulsion Laboratory / California Institute of Technology / NASA
2 Raytheon RIS
3 Oregon State University / Space Grant Consortium



1. Introduction

The EOS Microwave Limb Sounder (MLS) instrument produces over 9GB a day of both raw and processed atmospheric chemistry data. The MLS Science Computing Facility (SCF) houses resources that science team members use to analyze this data, specifically for performing tasks related to algorithm development, testing data processing configurations, analyzing instrument and science data, conducting validation studies, and developing/generating derived meteorological products. While the SCF's computing resources are adequate to perform science team members' tasks, there has been an on-going issue of managing them to perform optimally. Sun Grid Engine (SGE) is an open-source software package that is currently being tested at the SCF to help the facility better manage its resources and increase science team output by creating a grid network.

2. MLS SCF Problem Statement

This SCF sprawls across three locations:

1. Jet Propulsion Laboratory (Pasadena, CA)
2. Raytheon ITSS (Pasadena, CA)
3. New Mexico Tech (Socorro & Las Vegas, NM)

Available MLS SCF processing resources:

- ~764 CPUs
- CPU speeds ranging from 1GHz to 3.6GHz

Many resources are overtaxed while others sit idle

(Causes inefficiency and slows down processing and productivity)

3. Proposed Solution

- Grid technology has been proposed to help the MLS SCF better manage its resources
- A grid is a network that will connect various computer resources, that are disparate and geographically dispersed, to form a virtual organization^A
- Grids can be utilized to join desktops and clusters together to harness their combined processing power
- Grid networks coordinate job submission, use standard protocols, and deliver quality of service^B
- Grids also provide security, authenticate users, provide centralized resources, and prioritize jobs^C
- With grid technology, the SCF can become a more productive computing center for the geographically separated MLS team members
- Users can simply submit jobs and let the software choose where is best to run the job, regardless of its location
- Sun Grid Engine (SGE) provides an open-source package to connect computing resources. It allows for job prioritization, process monitoring, resource specification, and is PVM ready.

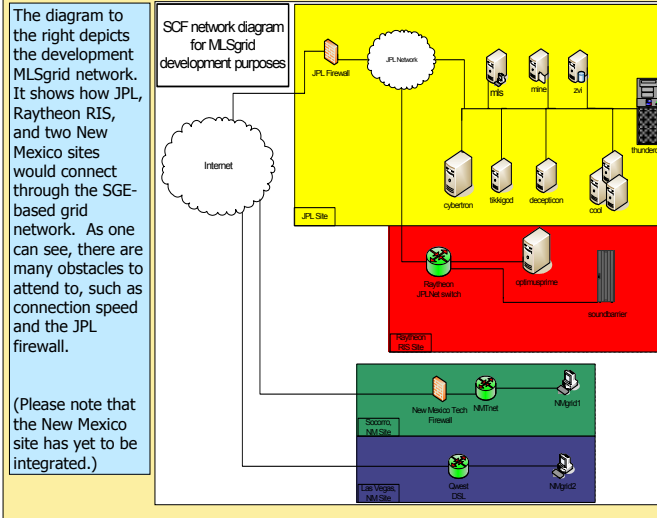
This grid will fully integrate:

- Desktops - Authentication
- Cluster computers - Databases
- File storage

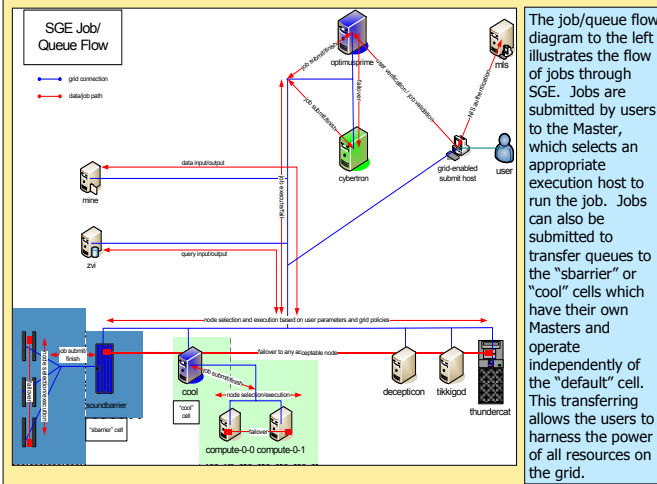
4. Available Development Resources

- optimusprime – master
- cybertron – shadow master
- mls – NIS server
- mine – file server
- thundercat – SUN station
- zvi – BerkDB server
- soundbarrier – 27-node cluster
- cool – 2-node cluster
- tikkigod – desktop/workstation
- deception – desktop/workstation

5. Development Grid Network Diagram



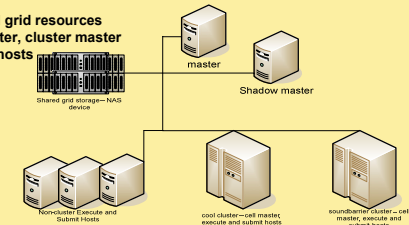
6. Grid Job / Queue Flow Under SGE



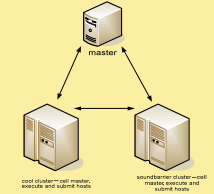
8. Sun Grid Engine (SGE) Implementation

1. Configure resources

- Install OS and implement shared grid resources
- Install SGE master, shadow master, cluster master
- Install SGE execute and submit hosts



2. Implement Transfer Queues Enabling communication between clusters

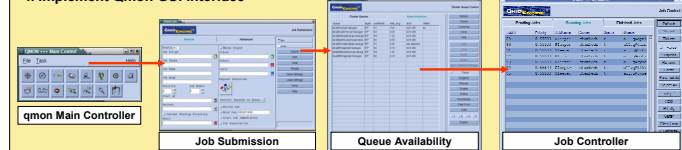


3. Implement Parallel Queues

- A special queue that separates the jobs into smaller portions for faster processing



4. Implement Qmon GUI interface



8. Current Status

- JPL and Raytheon grid configured
- Working to establish transfer queues
- Investigating alternative options including transfer queues over Globus

9. Benchmarks

MTOPS (Millions of theoretical operations per second) ^{D,E}	
Soundbarrier cell	→ 354,392
Cool cell	→ 7,000
Default cell	→ 40,141
Grid total	= 401,533

10. Future Work

- Investigate alternative methods of grid authentication – Globus, etc.
- Extend grid to additional SCF and remote resources in New Mexico
- Test grid functionality on real MLS science team jobs – stand-alone and parallel

References

- ^ASchoff, Jennifer M. (2002, August). Grids: The top 10 questions. *Scientific Programming, special issue on grid computing*, 10, no. 2, 103-111. Retrieved February 16, 2006, from <http://www.globus.org/alliance/publications/papers/topfen1.pdf>.
- ^BFoster, Ian. (2002, July 22). What is the grid? A three point checklist. *GridToday*. Retrieved February 15, 2006, from <http://www.gridtoday.com/02/07/22/101036.html>.
- ^CFoster, Ian, Carl Kesselman, and Steven Tuecke. (2001). The anatomy of the grid: Enabling scalable virtual organizations. *The International Journal of High Performance Computing Applications*, 15, no. 3, 200-222. Retrieved February 16, 2006, from <http://www.globus.org/alliance/publications/papers/anatomy.pdf>.
- ^DArchive Processor – CTP Calculations. (n.d.). Intel Corporation. Retrieved August 29, 2006, from <http://www.intel.com/support/processors/abios-020668.htm>.
- ^EInformation on how to calculate "composite theoretical performance" (CTP). (n.d.). The MOSIS Service. Retrieved August 8, 2006, from http://www.mosis.org/forms/mosis_forms/ECCN_CTP_Computation.pdf.