Grid Technology Provides a Cyber Infrastructure Applicable to NASA Applications

Thomas H. Hinke, Ph.D.

NASA Ames Research Center

Outline

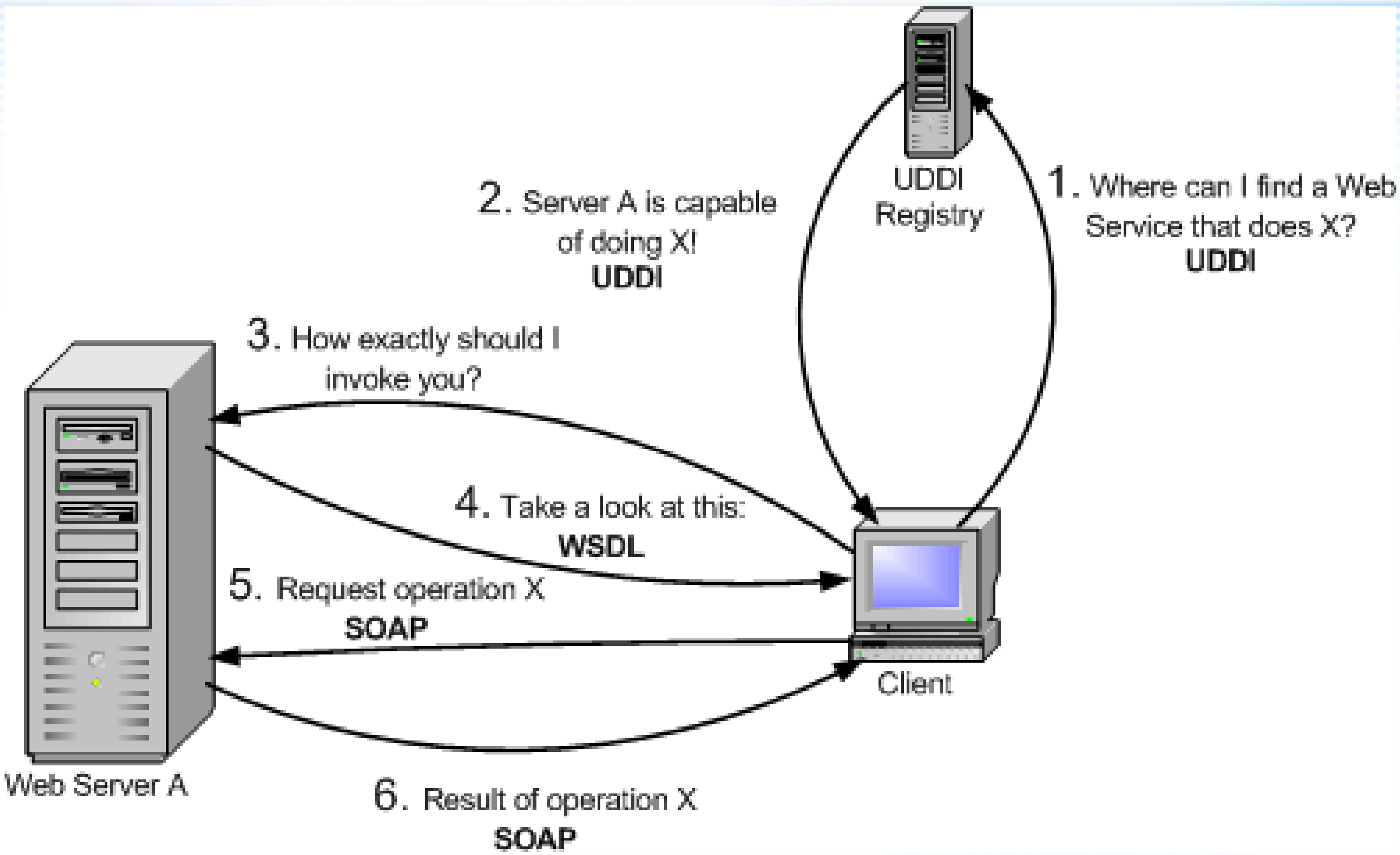
- What Grids provide and why are they useful to NASA
- NASA's role in grids and grid services
- Example applications

Grids are the Foundation for a Service-Based Infrastructure

Out of the commercial web/grid services community has come technology that

- Revolutionizes the ability to <u>easily</u> construct <u>extensible</u>
  - Mission control systems
  - Data handling systems
- Provides seamless access to various types of resources
  - Computational resources
  - Data archives
  - Scientific Sensors/Instruments
- Provides secure single-sign-on
- Provides an infrastructure that can support Grid services
  - Serve as building blocks for more complex services
  - Based on commercial web service technology

Web/Grid Services Support Software
Access To Services and Data



Graphic from Borja Sotomayor http://www.casa-sotomayor.net/gt3-tutorial/index.html

Grid Services Provide A Cyber Infrastructure for Applications

- Grid Services are based on Web services but support grid security
- They support remote access to data by software, not human access as is the goal of web page technology
    - Web has to do with human access to data using web browsers
    - Web services have to do with software access to data
- They are intended for loosely coupled distributed systems in contrast to CORBA and EJB (Enterprise Java Beans) which are intended for tightly coupled distributed systems
- Standardization efforts associated with Grid services
    - OGSA (Open Grid Services Architecture) by the Global Grid Forum
    - WS-RF (Web Service Resource Framework) by OASIS (Organization for the Advancement of Structured Information Standards) based on earlier work on Open Grid Services Infrastructure (OGSI) standard by the Global Grid Forum

Benefits of Grid Services

- Provided re-usable building blocks for other services and applications
- Reduced development costs for future space flight system development, replacement and upgrades
- Provides an architecture to allow the spiral development of complex, multi-center, multi-constellation mission management and operations environments
- Allows users and applications to easily use data and computational resources, irrespective of their location within the NASA community

Outline

- What Grids provide and why are they useful to NASA
- NASA's role in grids and grid services
- Example applications

Cyber Infrastructure

Vision

- To develop a grid services cyberinfrastructure that will provide *secure and seamless* access to a wide variety of space-oriented resources (e.g., computational, data, instruments and knowledge resources)
- To implement and facilitate the development of plug-n-play services focused on both core functionality and domain needs

Value Proposition

- A services-based architecture provides easy discovery, access and utilization of services thereby reducing the overall costs while increasing the efficiency, reliability and safety of space-oriented systems

Current Status of Grids at NASA

- NASA Ames Research Center (ARC)has developed the Information Power Grid that currently encompasses computational resources at
  - Ames Research Center, Glenn Research Center, Langley Research Center, JPL, with planning underway for Goddard Space Flight Center
- NASA is developing grid technology to increase the intelligence of the grid
- Since NASA may ultimately have many grids NASA ARC
  - Is providing technical support to other NASA and non-NASA organizations that are interested in deploying grid technology
  - Will deploy ARC-developed grid technology across the various NASA grids that may be developed

NASA Focus: Increasing Intelligence of Grid Processing and Data Handling

- Resource discovery service through brokers that select the “best” set of resources based on user requirements
- Execution manager service that can autonomously manage the user’s job as it moves through stages of grid processing
- Naturalization service that automatically tailors the processing environment on grid resources
- Dynamic access service that permits users to instantly, but with proper accountability, access needed computational resources across administrative boundaries, without having pre-established accounts on these machines
- Data discovery through distributed, grid-accessible metadata catalogs

NAS
NASA Advanced Supercomputing

NASA Focus: Increasing the Intelligence of Grid Management and Assessment

- Develop tools and processes that enable rapid integration of grid technologies for NASA applications
- Develop software that enables rapid installation, deployment, maintenance and monitoring of distributed services across heterogeneous and federated sets of resources
- Develop information services that accurately and quickly provide information about grid resources

NASA Grid Development

Application and User Oriented Services
- Portal Services
- Workflow Specification
- Query Structuring

Execution Management Services
- Workflow Management
- Resource Broker
- Naturalization Services
- Remote Execution
- Task Graph Execution

Data Management Services
- Replica Management
- Meta-data Catalogues
- Data Movement

Grid Management Services
- Distributed Information Management
- Dynamic User Access Management
- Event Management
- Test & Monitoring Service
- Information Service
- Security Infrastructure

Application Integration Processes and Tools

Compute & Data Resources

- ARC in-house R & D Efforts
- ARC partnership with external organizations
- External Technologies

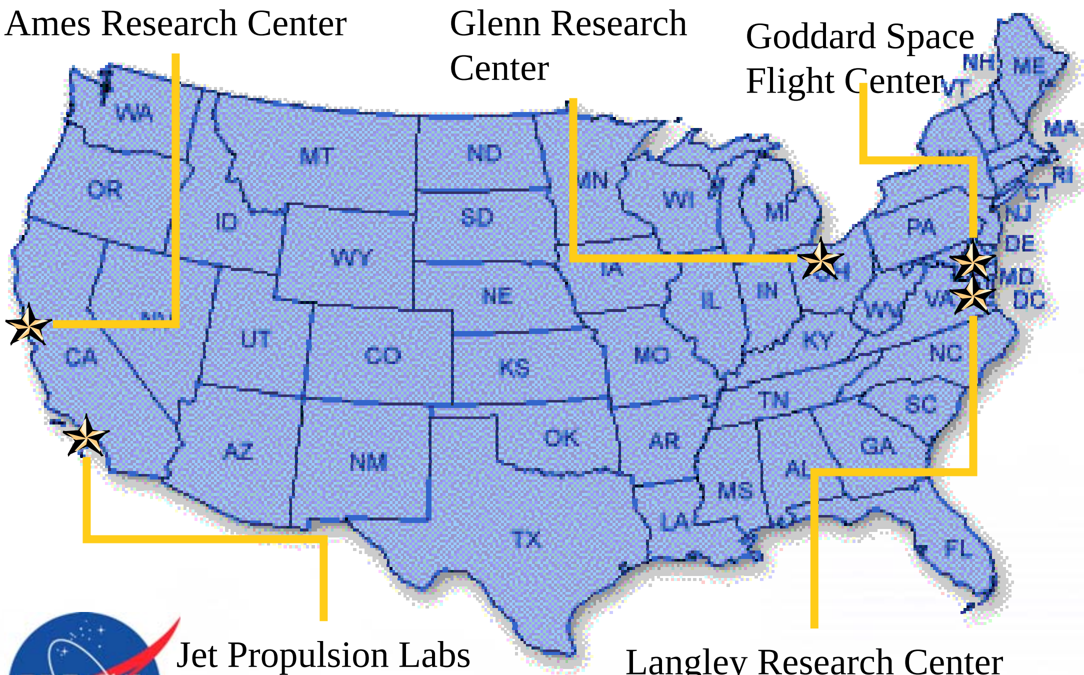
Information Power Grid

Vision:

To make the practice of large-scale science and engineering, as well as other widely distributed, data intensive NASA activities, much more effective than it is today.

Grid technology is the foundation to making this vision a success

- Grid Nodes
  1. 1024 CPU SGI O3K IRIX ARC
  2. 512 CPU SGI O3K IRIX ARC
  3. 512 CPU SGI Altrix LINUX ARC (in progress)
  4. 128 CPU (node) LINUX Cluster GRC
  5. 128 CPU SGI O2K IRIX ARC
  6. 128 CPU SGI O2K IRIX ARC
  7. 64 CPU SGI O2K IRIX ARC
  8. 32 CPU SGI O2K IRIX ARC
  9. 24 CPU SGI O3K IRIX ARC
  10. 24 CPU SGI O2K IRIX GRC
  11. 16 CPU SGI O2K IRIX LaRC
  12. 16 CPU SGI O2K IRIX ARC
  13. 12 CPU (node) LINUX Cluster (in progress)
  14. 8 CPU SGI O3K IRIX LaRC
  15. 8 CPU SUN Ultra SPARC3 ARC - Storage
  16. 8 CPU SGI O2K IRIX ARC - Storage
  17. 8 CPU SGI O3K IRIX JPL
  18. 8 CPU SGI O3K IRIX GSFC (planned)
  19. 4 CPU LINUX ARC (planned)



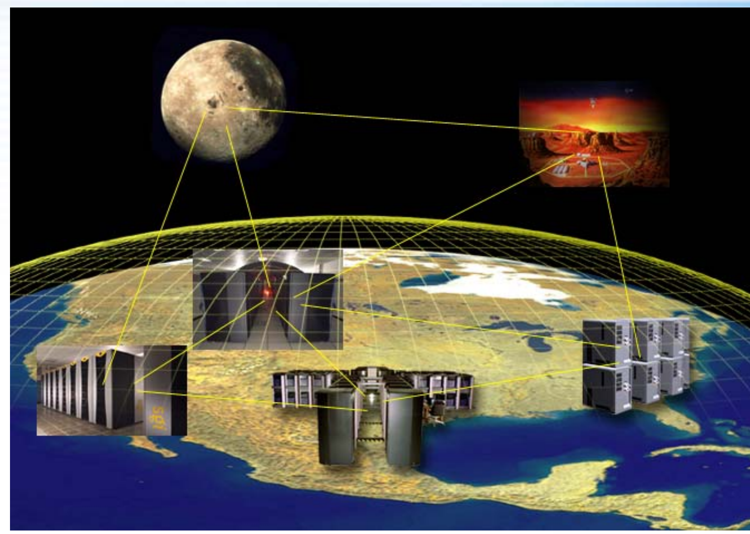
NASA

NAS
NASA ADVANCED SUPERCOMPUTING

Outline

- What Grids provide and why are they useful to NASA
- NASA's role in grids and grid services
- Example applications

Modeling and Simulation



Objective

- Enable NASA scientists, engineers, and managers to seamlessly use NASA's vast, distributed computational resources—from supercomputers to desktop workstations

Approach

- Develop/enhance grid service technologies: Certificate-based single sign-on, Broker, Execution Manager, Dynamic Access, Information Service

Benefits

- Transforms a set of distributed resources controlled by different administrative entities into a set of resources that can be seamlessly and easily used by all members of the NASA community
- More effective utilization of under-utilized resources, since the broker may locate resources not previously known to the user

Space Operations and Science Grid

Objectives

- Provide a service-based prototype to support distributed and secure access to Mission Control (MCC) and Payload Operations Integration Center (POIC) applications (In collaboration with MSFC and JSC)

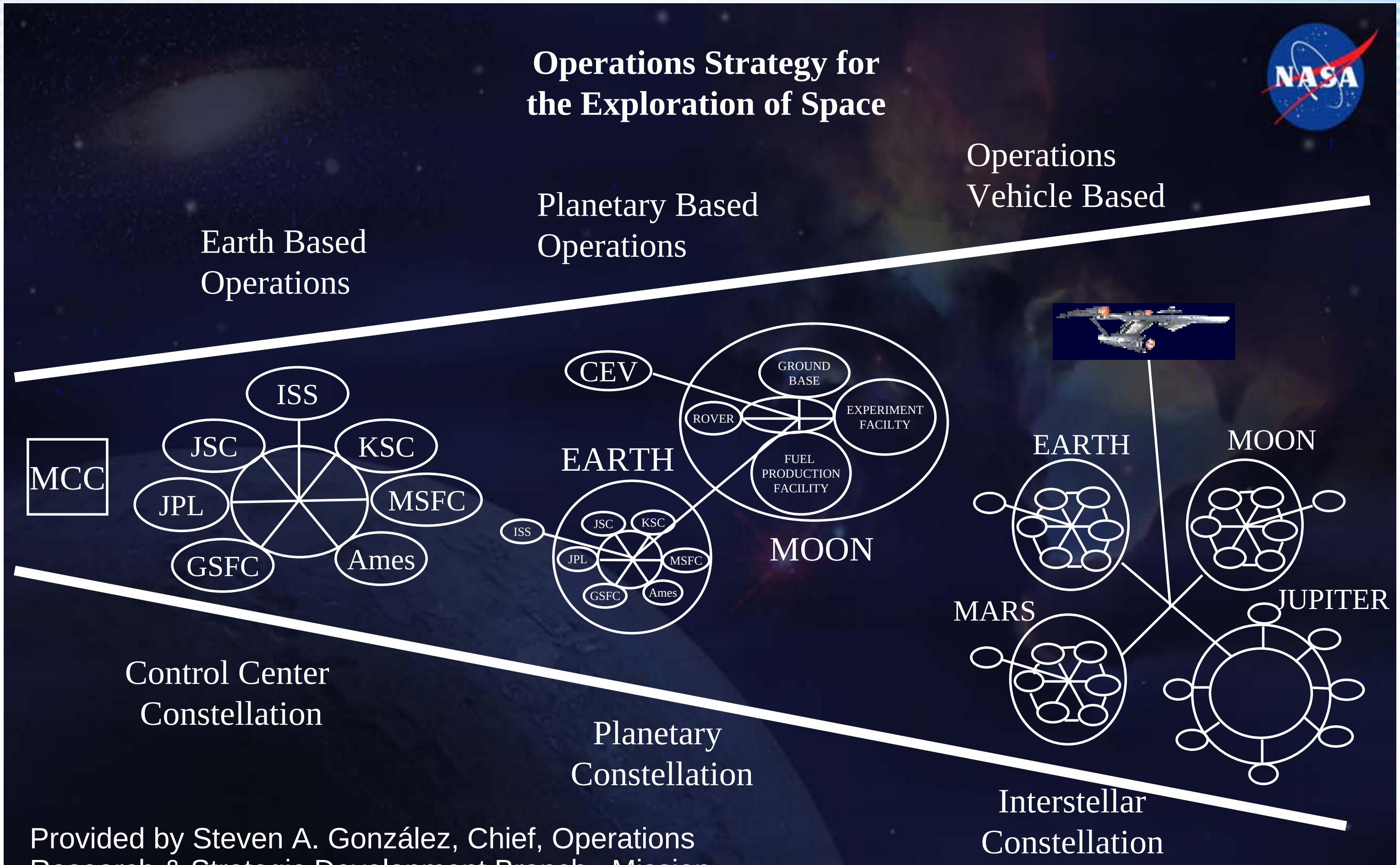
Approach

- Create a service-based infrastructure for Space Flight Operations from existing services in a non-operational shadow environment
- Transform existing applications into grid services such a payload telemetry processing, caution and warning system and health telemetry
- Develop portals providing access to community-specific services for end-user communities such as Payload Principal Investigators, Payload Control Center, Flight Control Center, Educational Outreach

Benefits

- Reduce development costs for future space flight system development, replacement and upgrades
- Provide a single interface for secure and distributed access to new and current MCC & POIC services at significantly lower costs

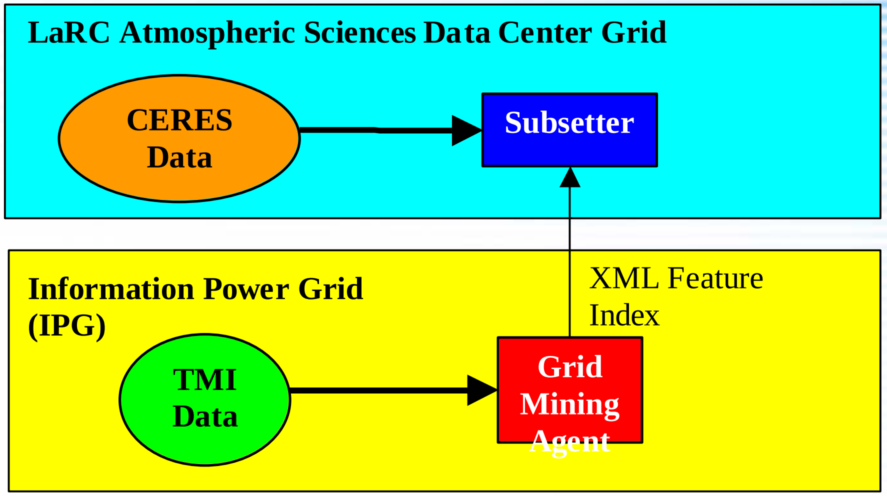
Mission Control Center Evolution



Provided by Steven A. González, Chief, Operations Research & Strategic Development Branch, Mission Operations Directorate, Johnson Space Center

Distributed Generation of Earth Science Data Products Using Grids

Objectives: Demonstrate the use of grid technology to support intra-center and inter-center generation of Earth Science data products



Approach:

- Used grid technology to develop subsetter for CERES data at LaRC Atmospheric Sciences Data Center
- Used Grid Miner to mine GSFC TRMM/TMI data (that was cached at on ARC mass storage system) for mesoscale convective systems (MCS)
- Used grid-enabled Storage Resource broker to extract data directly from mass storage system for use by miner
- Used grid to transfer MCS feature index to LaRC to subset CERES data

Benefit: Grid technology eased the effort to generate a data product that involved data from two NASA centers

Committee on Earth Observation Satellites (CEOS) Grid Testbed

Objectives: The CEOS GRID Task Team will develop technical requirements and identify GRID technologies and services to be implemented in testbed locations

Approach:

- ARC and others supported CEOS Grid Workshop in April 2002
- CEOS Grid Testbed participants began work in 2002 and have grown since then
- ARC provided grid consulting as well as host and user grid certificates

Participants:

- EOSDIS & George Mason University (GMU)
- European Space Agency (ESA)
- DutchSpace
- NOAA Operational Model Archive & Distribution System (NOMADS)
- University of Alabama – Huntsville (UAH)
- United States Geological Survey – EROS Data Center (EDC)
- NASA Advanced Data Grid (ADG)
- China Spatial Information Grid (SIG)
- ARC (supplying host and user certificates and grid consulting)

- Various CEOS participants are developing grid applications

Benefits: Vehicle for exploring the advantage of Grid technology for Earth Science

Summary of Selected CEOS Projects

- USGS (US Geological Survey):
  - Explore use of GRID technologies for the delivery/reception of earth science data
- NOAA (National Oceanic and Atmospheric Administration):
  - Provide access to climate and numerical weather prediction models for analysis and intercomparison
  - Foster research to study complex earth systems using collections of distributed data
- ESA (European Space Agency):
  - Generic infrastructure to allow seamless plug-in of specific data handling & application services
  - Support on-demand user-driven data integration
- NASA GSFC (Goddard Space Flight Center) and GMU (George Mason University):
  - Integration of Grid and Open GIS Consortium (OGC) Web Services
- GMU (George Mason University):
  - Provide the ability to advertise and deliver virtual datasets
- UAH (U. of Alabama in Huntsville):
  - Compute-intensive data mining and machine learning applications in the Earth sciences
- DutchSpace and ESA/ESRIN (European Space Research Institute):
  - Simulators of EO instruments and data processing software working together using Computational Grid technology

NASA ADVANCED SUPERCOMPUTING

Questions?