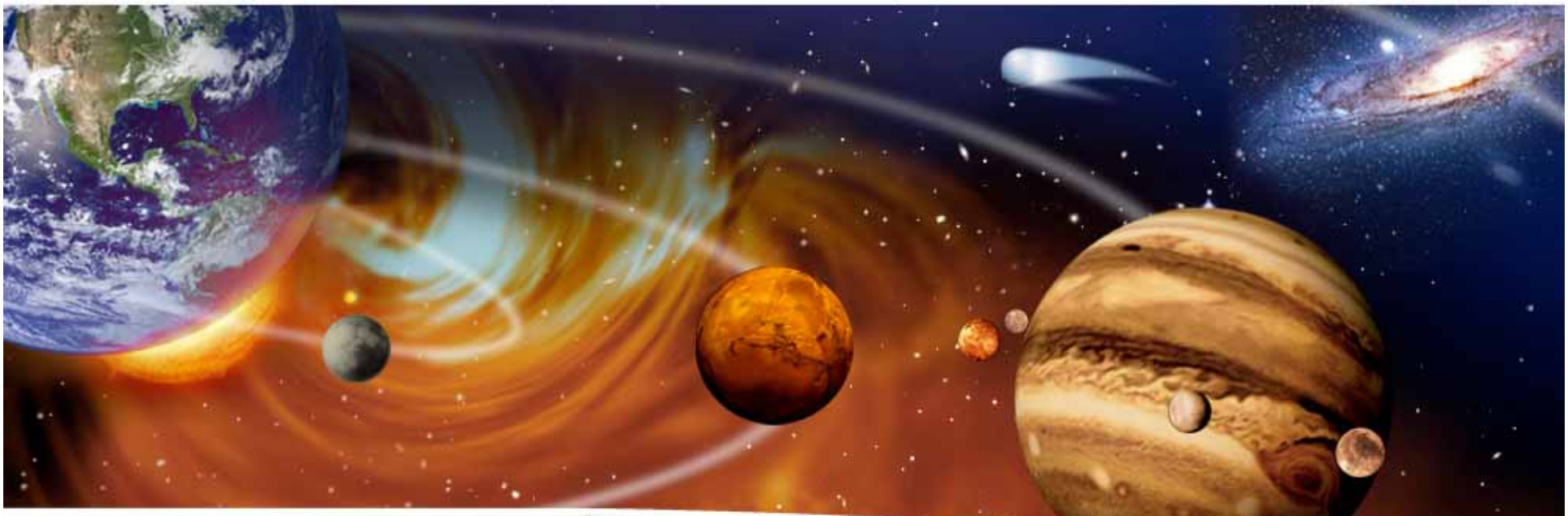


Project Columbia

Tsengdar Lee

Science Mission Directorate

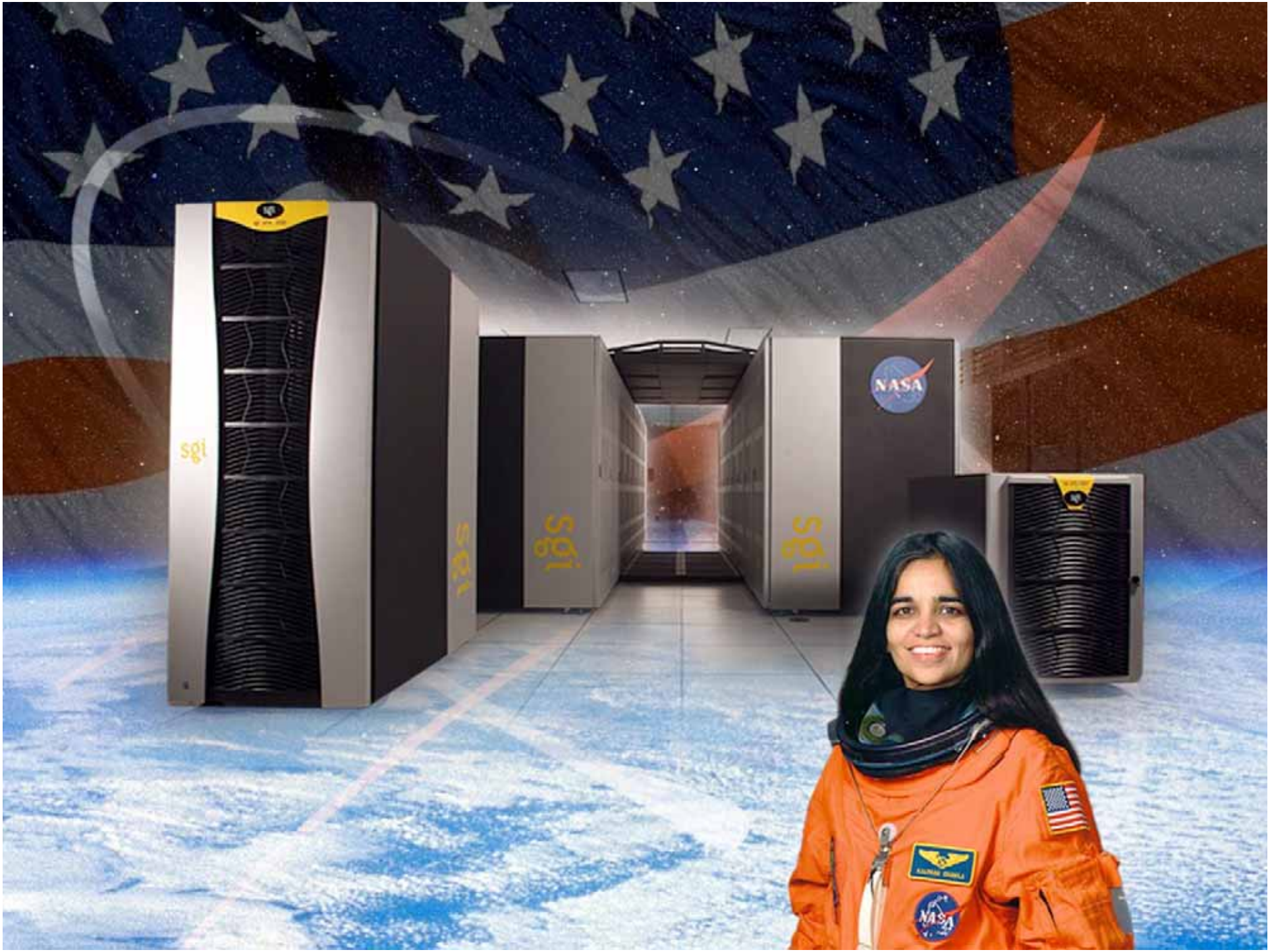
November 1, 2004

Thrust of the Talk

- Project Columbia history and current status
- What are we doing on Project Columbia
  - Some significant results and new challenges
- What are we planning to do with Project Columbia

Project Columbia History

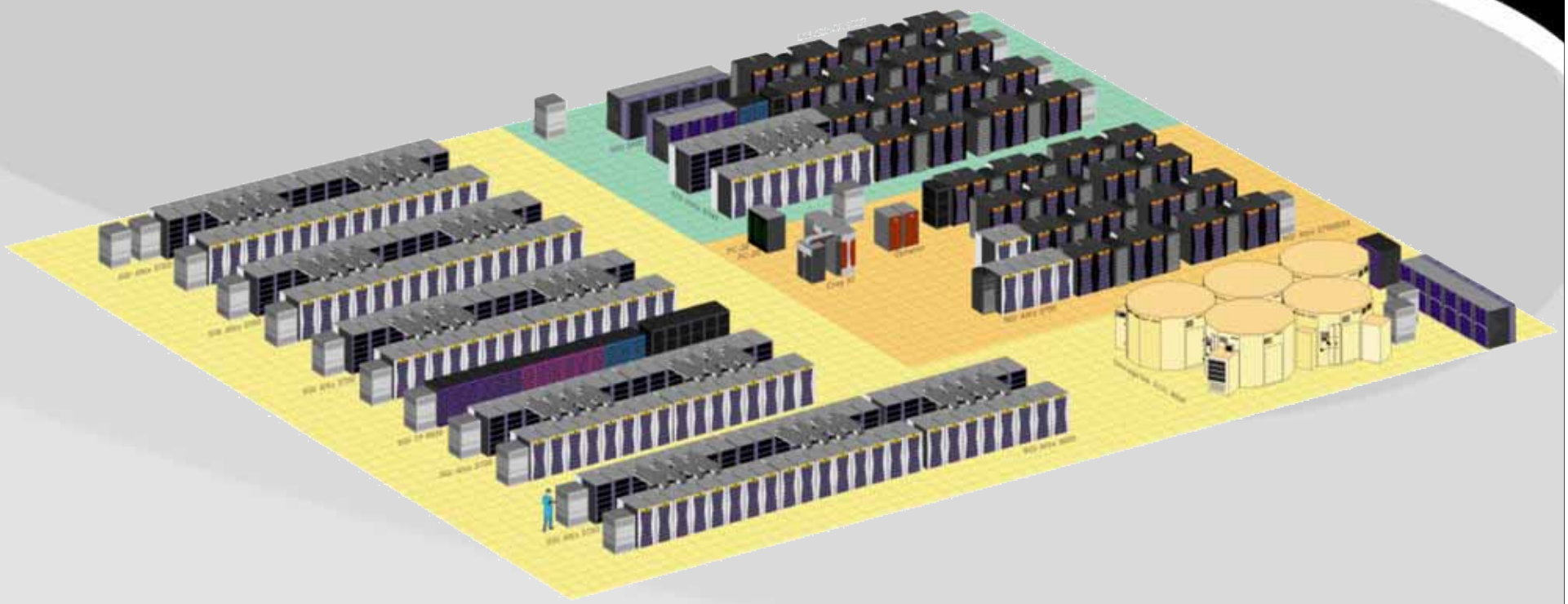
- Started in August 2003 as an Earth Science Enterprise and Aeronautics Technology Enterprise collaboration
- Demonstrated Earth science modeling project (ECCO) at SC2003 Conference
- The first SGI Altix system dedicated to ‘Kalpana’ in April 2004
- Initiated partnership discussions with Intel and SGI in April 2004
- ARC proposed to NASA Executive Committee in mid-May 2004
- Approved by OMB and Congress in June 2004
- Funded in June 2004 by NASA
- System delivery started in early July
- Installation completed mid-October
- Achieved 42.7 Teraflops sustained LinPack benchmark with 16 of the 20 systems
- Currently world’s #1 supercomputer


sgi
sgi
sgi
sgi
NASA

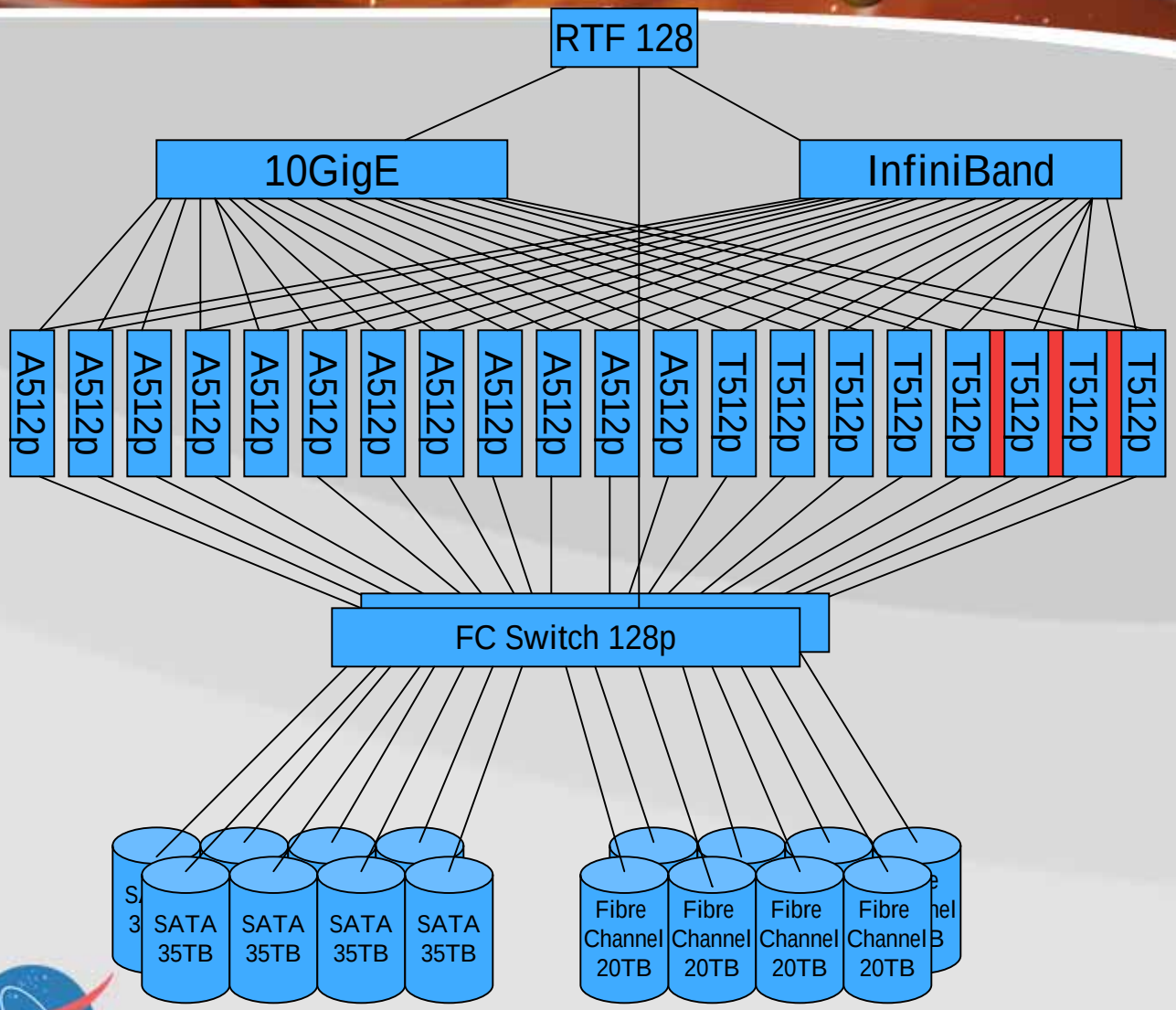
Project Columbia

- Partnership between NASA and industry to significantly enhance the national compute capability
- Unique opportunity that met both industry and NASA objectives
- In line with HECRTF findings
- National asset available to multiple agencies through competitive selection process
- Unique capability built from proven technologies
- Asset enhances capability that was not currently being targeted by other leadership class systems
- Immediate capabilities provide NASA with increasing capabilities
- Includes networking component to connect NASA centers for scientific research

Columbia Target



Components



Front End
- 128p Altix 3700 (RTF)

Networking
- InfiniBand Switch (288port)
- InfiniBand (4 per 3700, 8 per 3700Bx2))
- 10GigE Switch 32-port
- 10GigE Cards (1 Per 512p)
- GigE Switch 384-port, 96-port
- GigE (12 per 512)
- Altix 3700Bx2 2048 Numalink Kit

Compute Nodes
- Altix 3700 12x512p
- Altix 3700Bx2 8x512p

Storage Area Network
-Brocade Switch 2x128port

Storage (440 TB)
-FC RAID 8x20 TB (8 Racks)
-SATARAID 8x35TB (8 Racks)

Mass Storage (17 PB)
-17 PB MSS

Current Achievements

- Installation completed in mid-October
- Origin reconfiguration, decommissioning and movement complete
- Infrastructure
  - Power Dist. Units shipments received and installed
  - New Cooling Loops Installed
  - New Fiber Optics Backbone
  - Cooling Systems tested
  - New Wiring and Distribution Panels Completed
- Storage
  - Disk farm in place 400TBytes
  - Silos Reconfigured
- Applications
  - Significant progress on fvGCM and ECCO
  - Significant current usage by Space weather

Speed Racer

NASA's Columbia machine is a top contender for the title of world's fastest supercomputer

Building Blocks

- Each cabinet of the Columbia supercomputer, built by SGI for NASA, has 32 processors. The system, unveiled Oct. 26 at the Ames Research Center in Moffett Field, Calif., has 10,240 Itanium 2 processors in total.
- (Photo credit: Stephen Shankland)

Internetworking

- Massive white "NUMAlink" cables packed with copper wires connect 16 cabinets into a single 512-processor computer. Twenty of these Silicon Graphics Inc. machines, connected with InfiniBand networking equipment from Voltaire, make up NASA's Columbia supercomputer.

- (Photo credit: Stephen Shankland)

New Packaging

- SGI, showing off the components of NASA's Columbia supercomputer, gave a glimpse of its yet-to-be-announced BX2 model of the Altix 3700 server.
- (Photo credit: Stephen Shankland)

Storage System

- The storage system of the Columbia supercomputer, built by SGI for NASA, holds 440 terabytes of data.
- (Photo credit: Stephen Shankland)

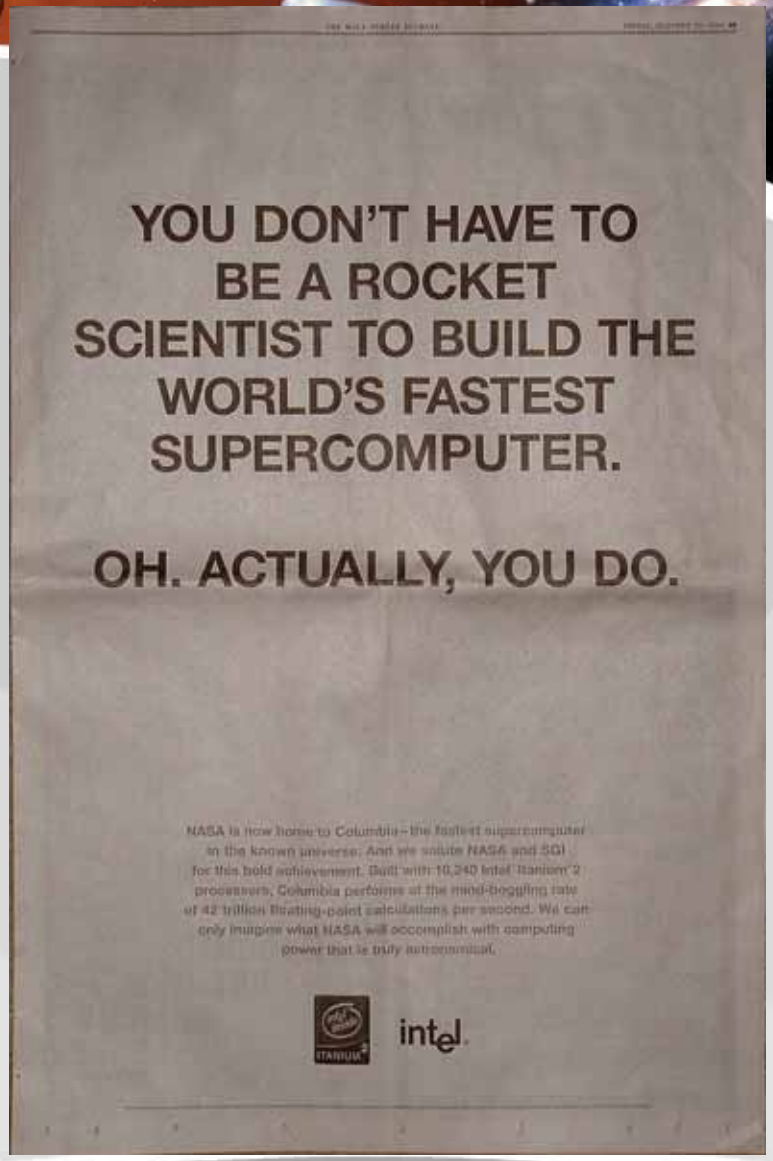
Cooling System

- The forthcoming BX2 model of SGI's Altix 3700 packs 64 Itanium 2 processors into one cabinet, compared to 32 for its predecessor, but the resulting heat led SGI to offer the system with liquid cooling built into the back door--an option NASA chose for its Columbia supercomputer.

- (Photo credit: Stephen Shankland)

WSJ 10/29/2004 A5

- "NASA is now home to Columbia – the fastest supercomputer in the known universe. And we salute NASA and SGI for this bold achievement. Built with 10,240 Intel Itanium® 2 processors, Columbia performs at the mind-boggling rate of 42 trillion floating-point calculations per second. We can only imagine what NASA will accomplish with computing power that is truly astronomical. - Intel"



Challenge Remains

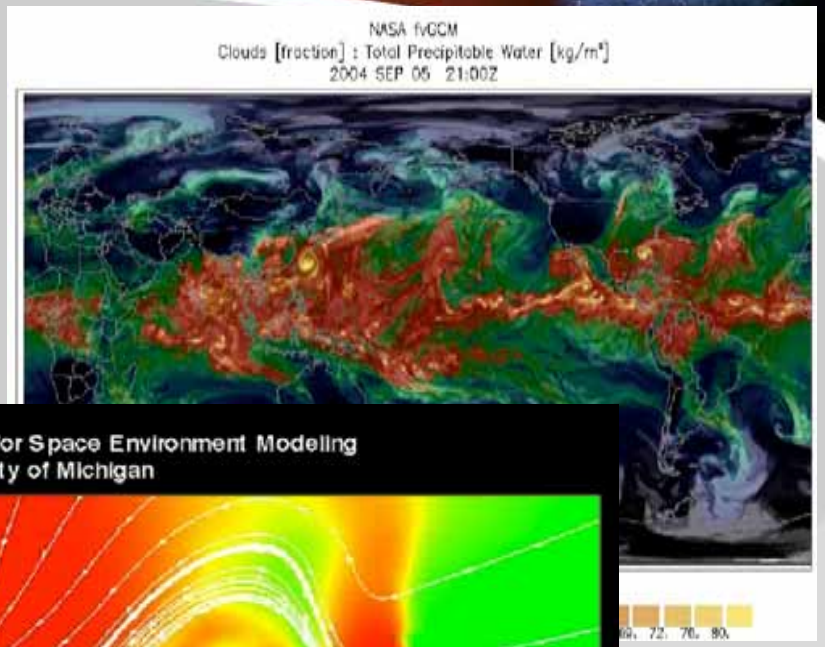
- Project Columbia is funded by all mission directorates
- It is funded to FY07
- No funding beyond FY07 in the current plan
- Agency is pushing a “service pool” model for the future funding
- Science Mission needs to articulate and advocate the “need”

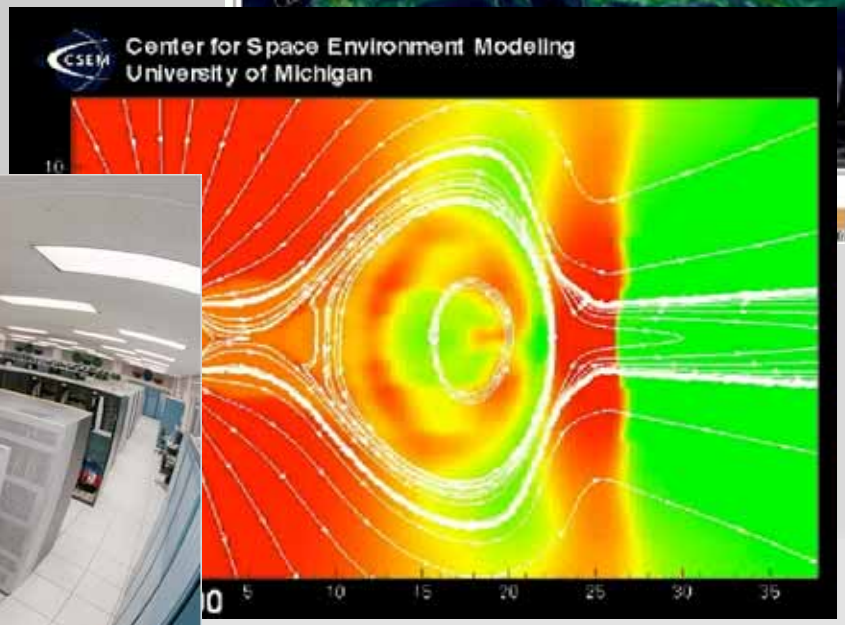
Thrust of the Talk

- Project Columbia history and current status
- What are we doing on Project Columbia
  - Some significant results and new challenges
- What are we planning to do with Project Columbia

Project Columbia Supports Earth-Sun Modeling and Analysis

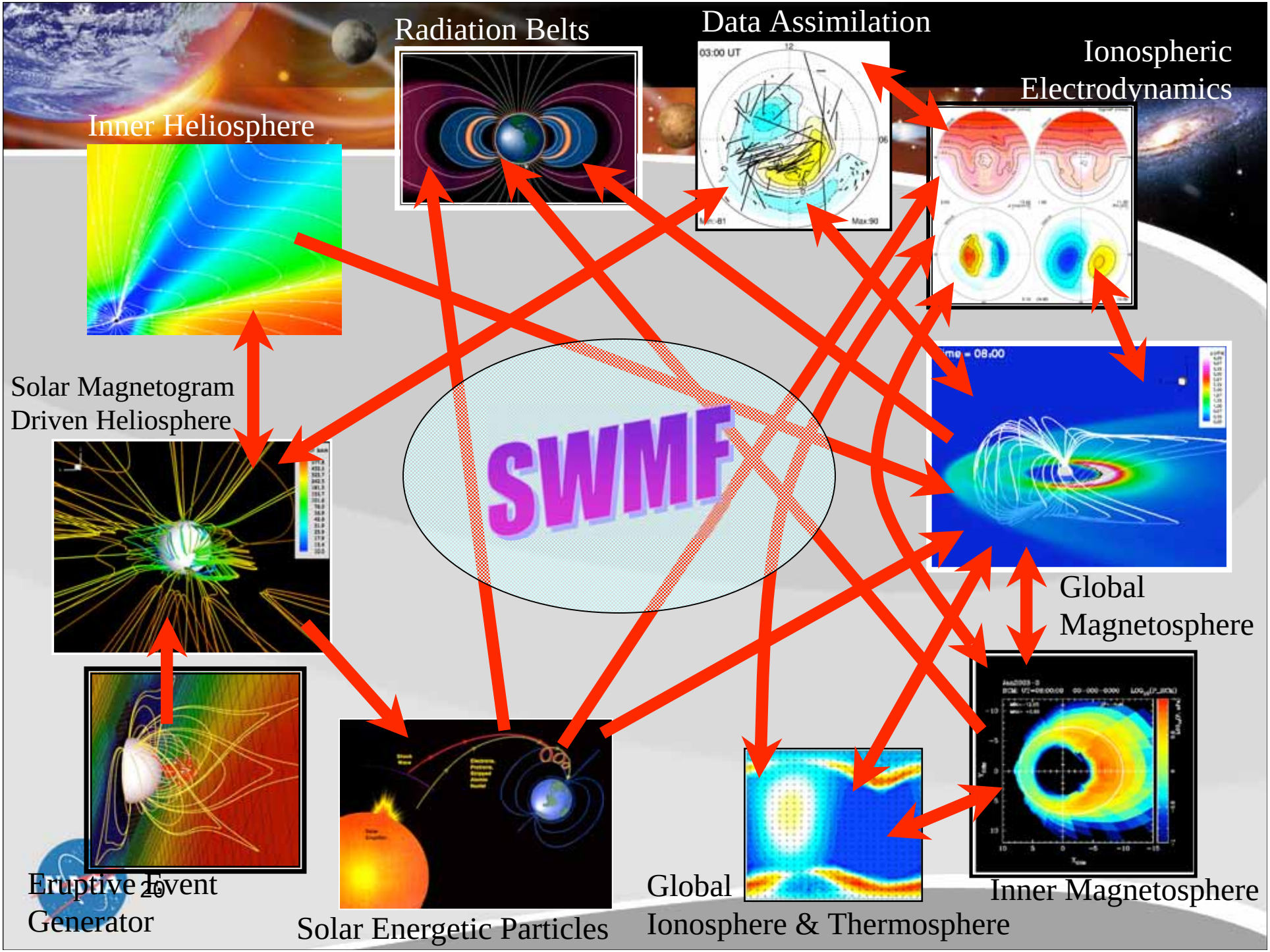
- Project Columbia dedicated 10/26
- World's fastest computer with 42.7 Teraflops throughput
- 28% allocated to SMD
- Earth-Sun modeling has been the prime usage of the systems
- Demand exceeds supply





Successful computing

- SWMF
  - SWMF has been developed at the University of Michigan under the NASA Earth Science Technology Office (ESTO) Computational Technologies (CT) Project to provide “plug and play” Sun-to-Earth simulation capabilities to the space physics modeling community
    - SWMF has successfully migrated to the Project Columbia
- ECCO
  - Continued success in ocean modeling has improved the model and the work was able to continue during the very busy RTF activities
- fvGCM
  - Because of the very promising results from the 1/4° fvGCM runs, GSFC is experimenting numerical weather predictions (NWP) in real time during the hurricane seasons
    - Goals are to predict hurricanes accurately in advance and delivery NWP products in time


Radiation Belts
Data Assimilation
Ionospheric Electrodynamics
Inner Heliosphere
Solar Magnetogram Driven Heliosphere
SWMF
Global Magnetosphere
Eruptive Event Generator
Solar Energetic Particles
Global Ionosphere & Thermosphere
Inner Magnetosphere

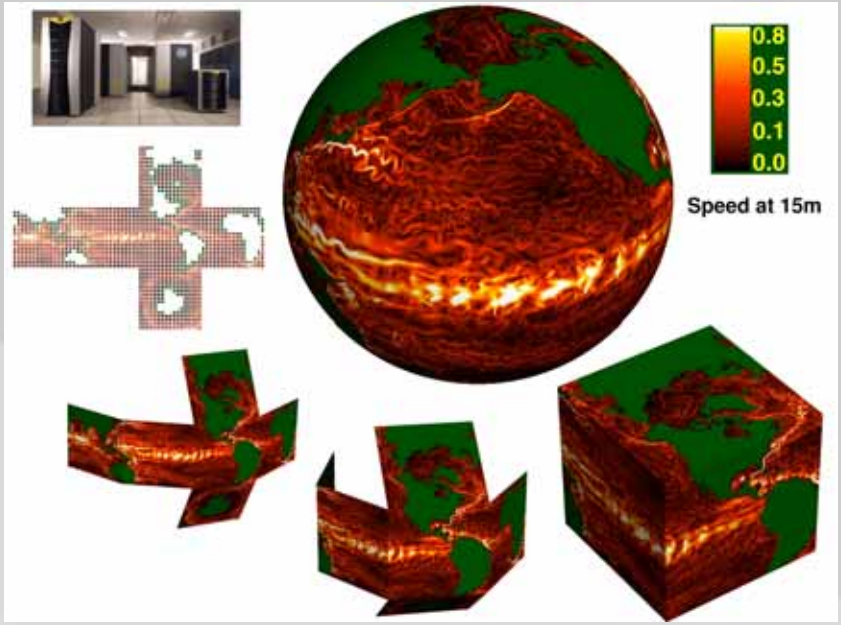
Space Weather Modeling Framework (SWMF)

- SWMF comprises a series of interoperating models of physics domains, ranging from the surface of the Sun to the upper atmosphere of the Earth.
- Currently, the SWMF links together nine models:
  - Magnetogram driven Global Solar Corona
  - Solar Eruption Generator
  - Inner Heliosphere
  - Solar Energetic Particles
  - Global Magnetosphere
  - Inner Magnetosphere
  - Radiation Belts
  - Ionosphere Electrodynamics
  - Upper Atmosphere

Estimating the Circulation and Climate of Ocean (ECCO)

- Estimating the circulation and climate of the ocean
- 1/6-degree lat-lon grid: 1440 x 1088 x 50
- “Cube-sphere” grid: 6 * 510 x 510 x 50
- ~78 million points
- 10 minute timestep
- Decadal simulations
- ~0.5 TB/day
- Large volume of ocean observations (altimetry, in situ temperature profile, SST, in situ sea level and bottom pressure, historical hydrographic data)
- Large volume of atmospheric forcing fields (NCAR/NCEP reanalysis, NCEP global data assimilation, QuikScat winds)

Distributing ECCO Results



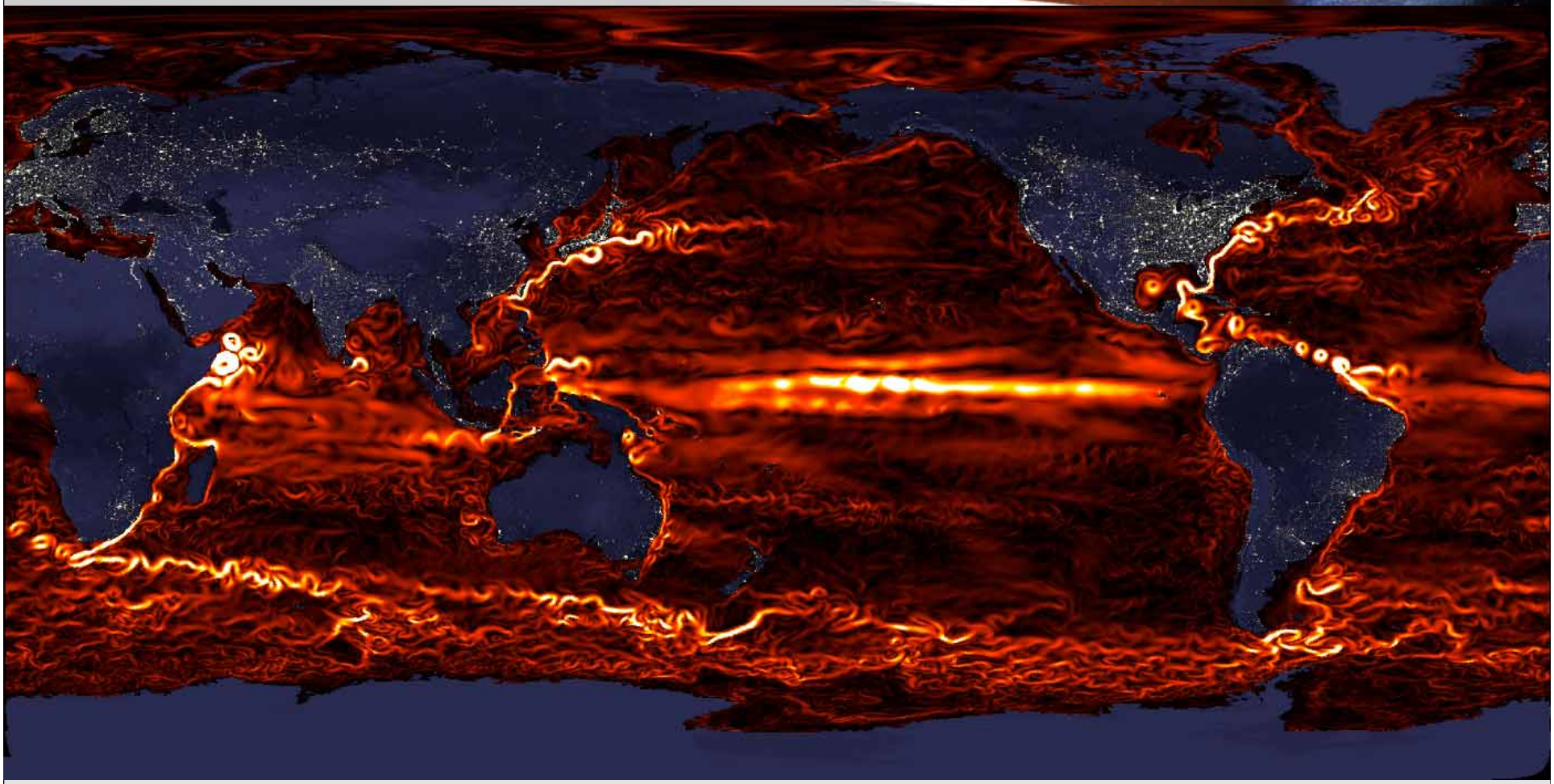
ECCO cube-sphere model domain comprises 78,030,000 grid points.

Depending on what diagnostics are saved, each 10-year integration can generate 5 Tbytes or more.

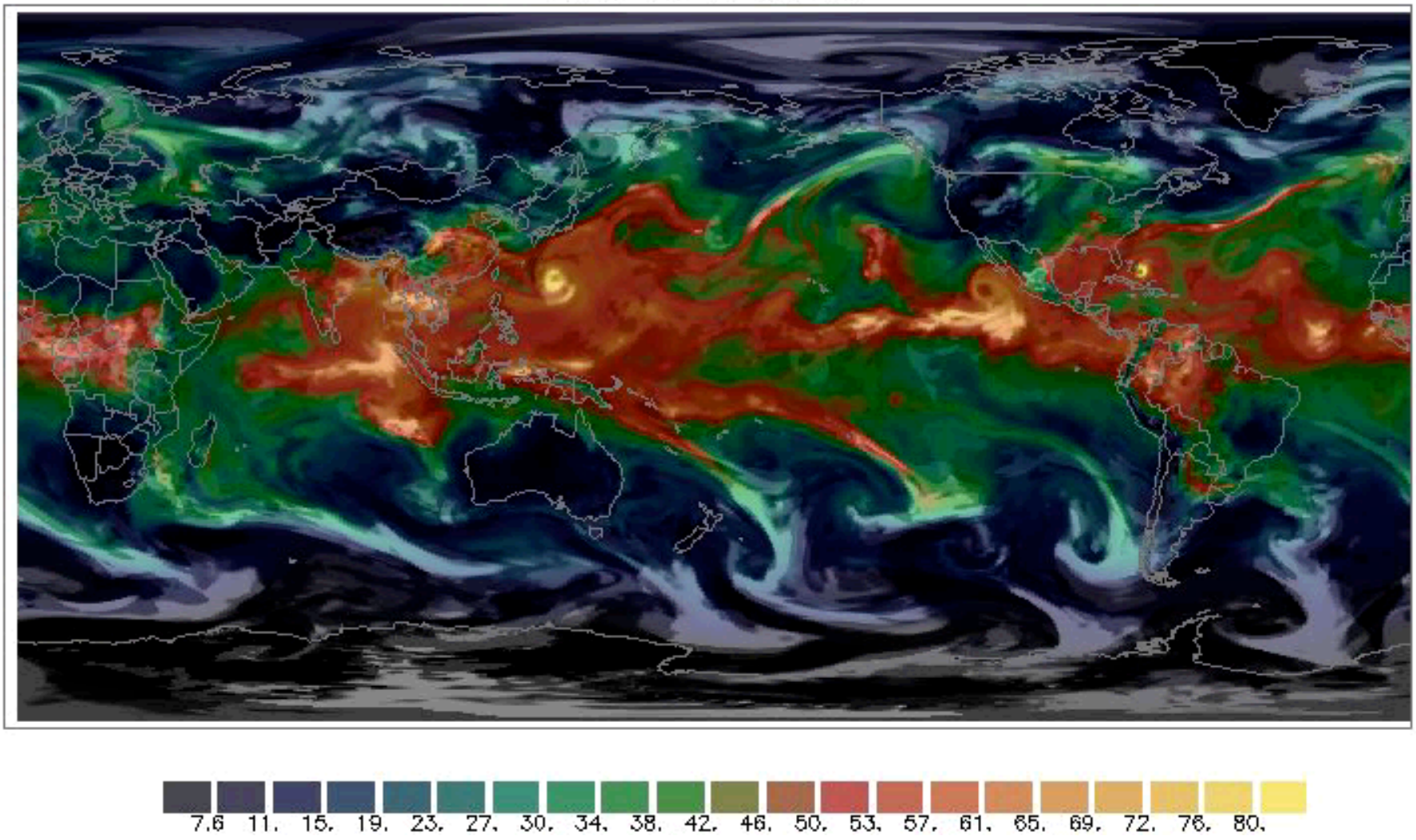
Requirements:

1. Fast network connection, order 10 Mbyte/s or greater to various hubs, e.g., Colorado, Goddard, Hamburg, MIT, SCRIPPS, Seattle
2. Mechanism for filtering/subsampling ocean state estimates prior to transfer, e.g., a server and a tape silo accessible to outside users

ECCO Data Visualization



fvGCM Hurricane Forecast

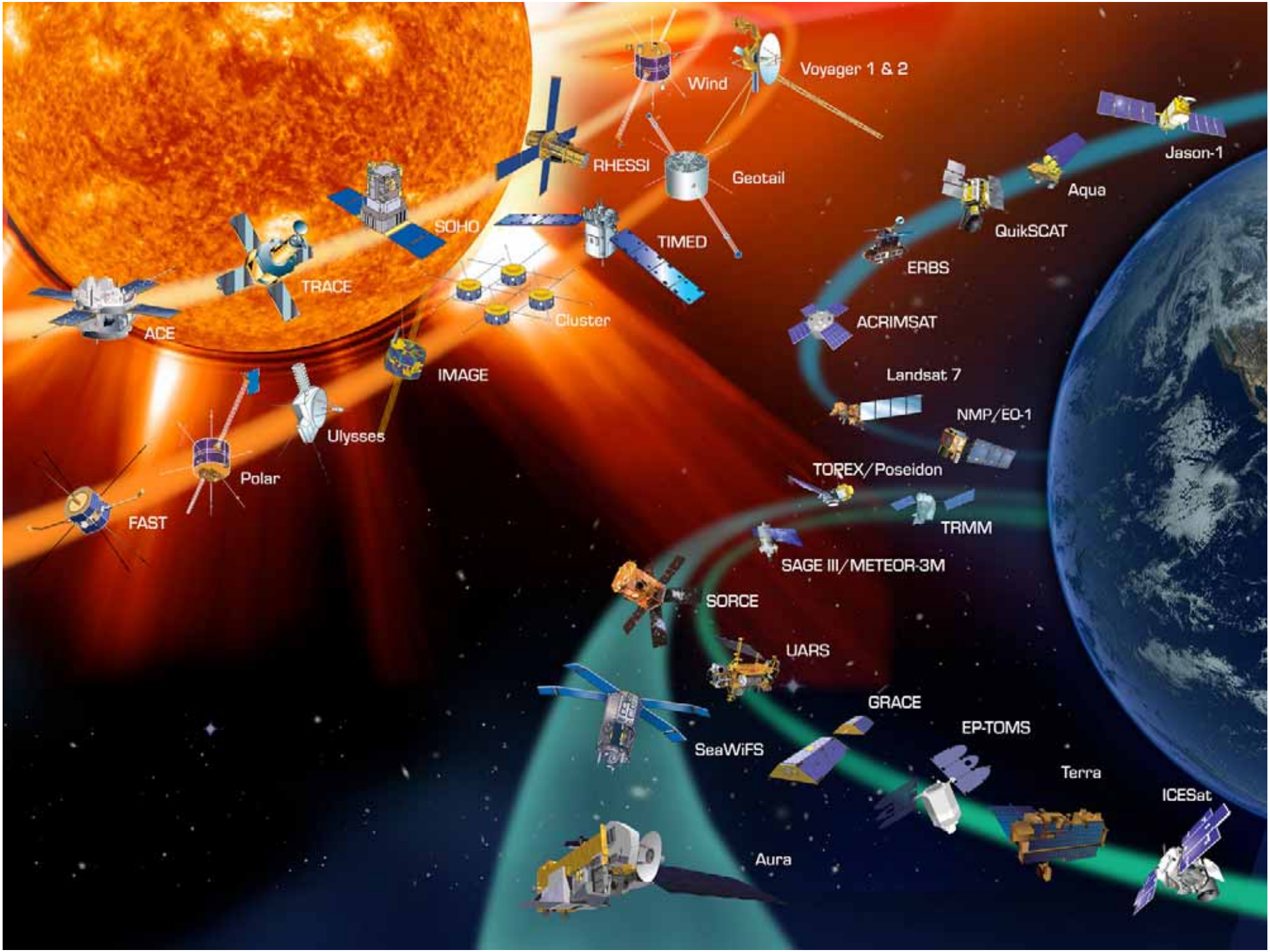


Finite Volume General Circulation Model (fvGCM)

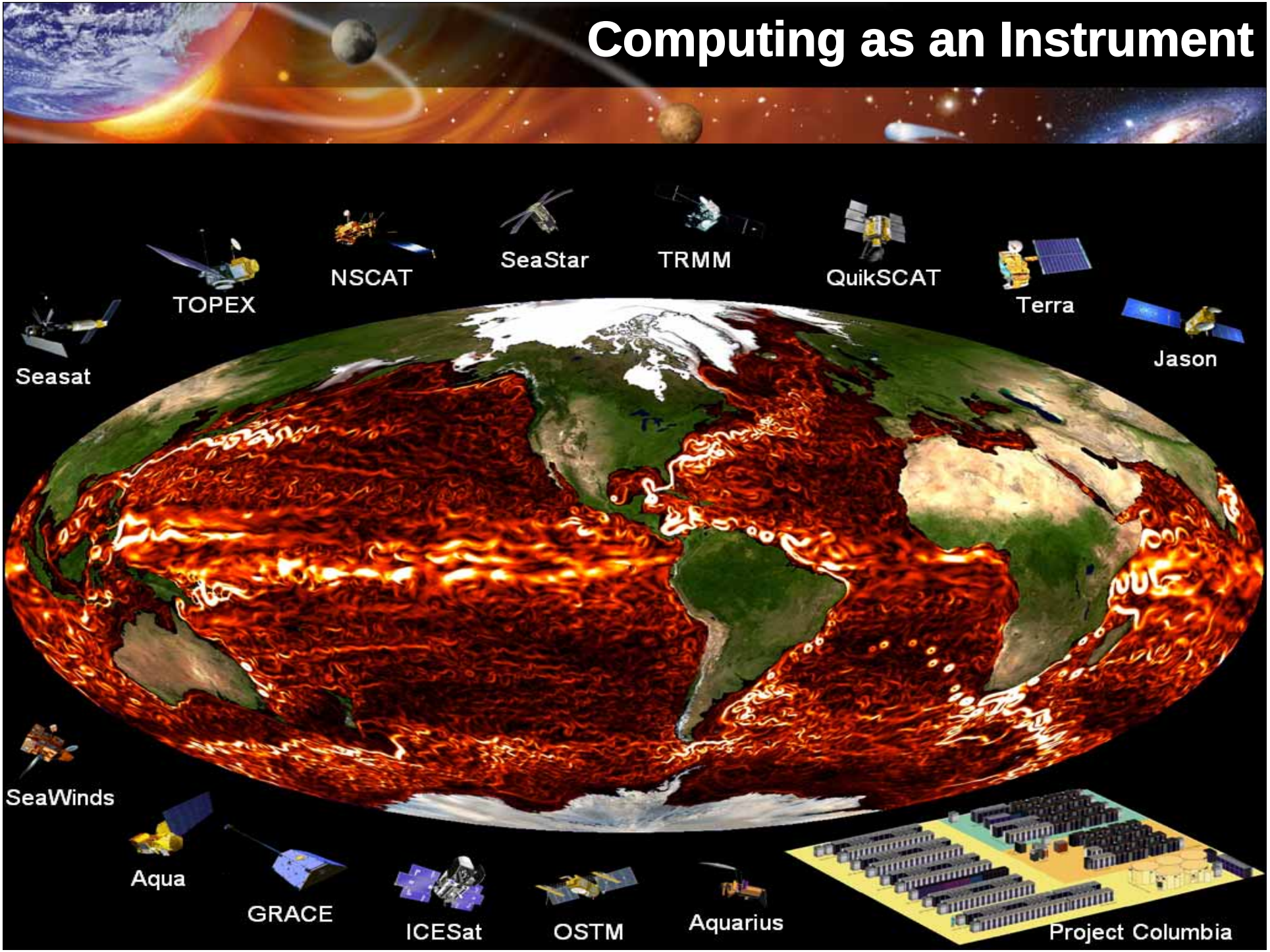
- Finite volume general circulation model
- 1/4-degree global grid: 1000 x 721 x 32
- 23 million points
- 15 minute timestep for physics
- 45 seconds timestep for dynamics
- ~1 TB data set for 5-day forecast
- Twice daily of 200MB initial condition transferred from NCEP to Project Columbia
- Output data transferred to GSFC for detail analysis
- Security concerns have significantly slowed the data transfer
- Initial work under Project Columbia has resulted in a .25 deg version of the FVGCM that is demonstrating a new capability to represent tropical convective clusters in a global atmospheric model and dramatic improvements in the simulation (prediction) of hurricane track and intensity

Thrust of the Talk

- Project Columbia history and current status
- What are we doing on Project Columbia
  - Some significant results and new challenges
- What are we planning to do with Project Columbia

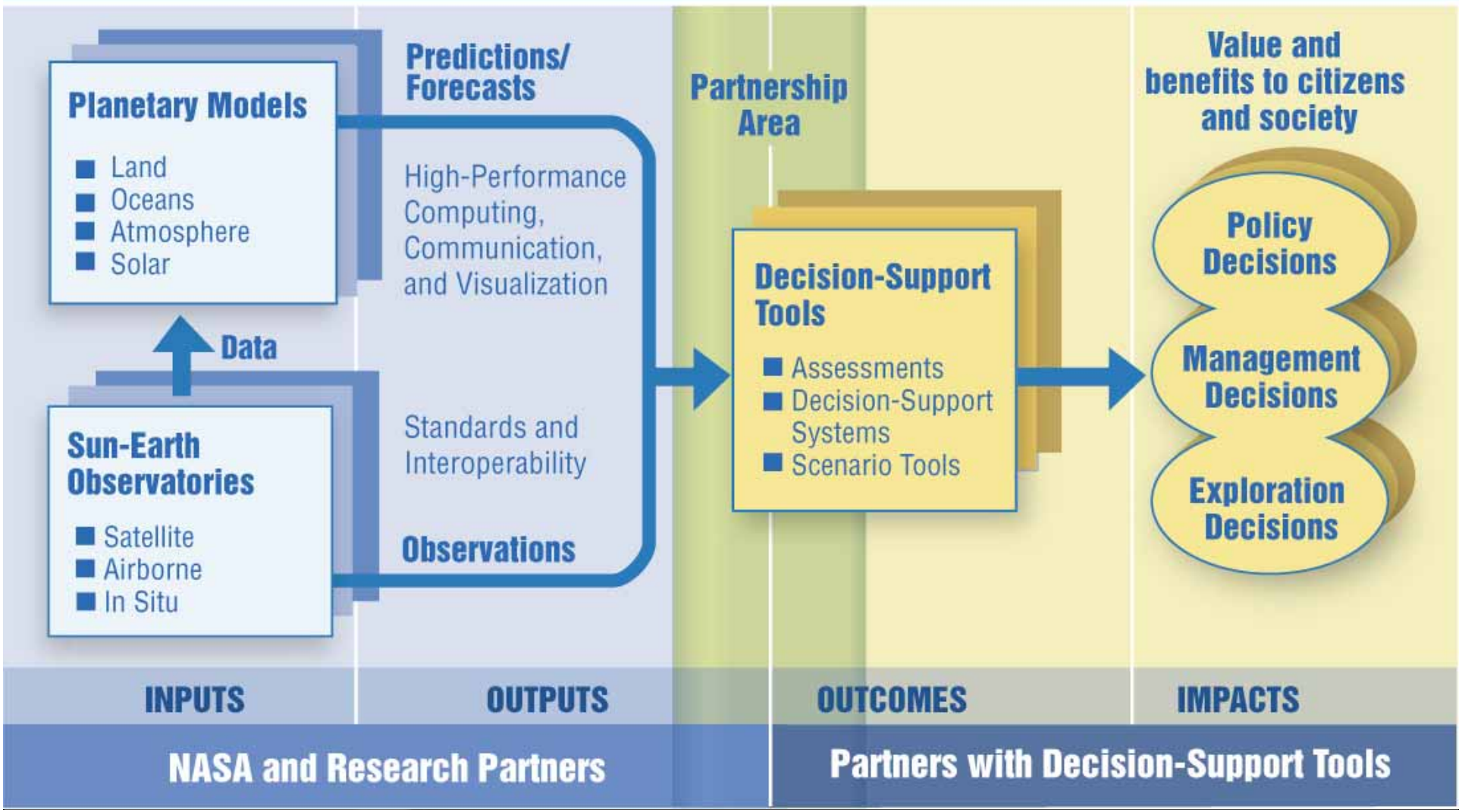

Wind
Voyager 1 & 2
RHESSI
Geotail
SOHO
TIMED
TRACE
Cluster
ACE
IMAGE
Ulysses
Polar
FAST
Jason-1
Aqua
QuikSCAT
ERBS
ACRIMSAT
Landsat 7
NMP/EO-1
TOREX/Poseidon
TRMM
SAGE III/METEOR-3M
SORCE
UARS
GRACE
EP-TOMS
SeaWiFS
Terra
ICESat
Aura

Computing as an Instrument



System of Systems Framework

Applied Sciences Program Approach to Integrated System Solutions



Data Acquisition to Data Access

Data Acquisition

Flight Operations, Data Capture, Initial Processing & Backup Archive

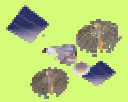
Data Transport to DAACs

Science Data Processing, Data Mgmt., Data Archive & Distribution

Distribution, Access, Interoperability & Reuse

Spacecraft

Tracking & Data Relay Satellite (TDRS)



Ground Stations

Polar Ground Stations

Data Processing & Mission Control

NASA Integrated Services Network (NISN) Mission Services

EOSDIS Science Data Systems (DAACs)

REASoNs

WWW IP Internet

Science Teams

Measurement Teams

Research

Education

Value-Added Providers

Interagency Data Centers

International Partners

Use in Earth System Models

Benchmarking DSS

TECHNOLOGY

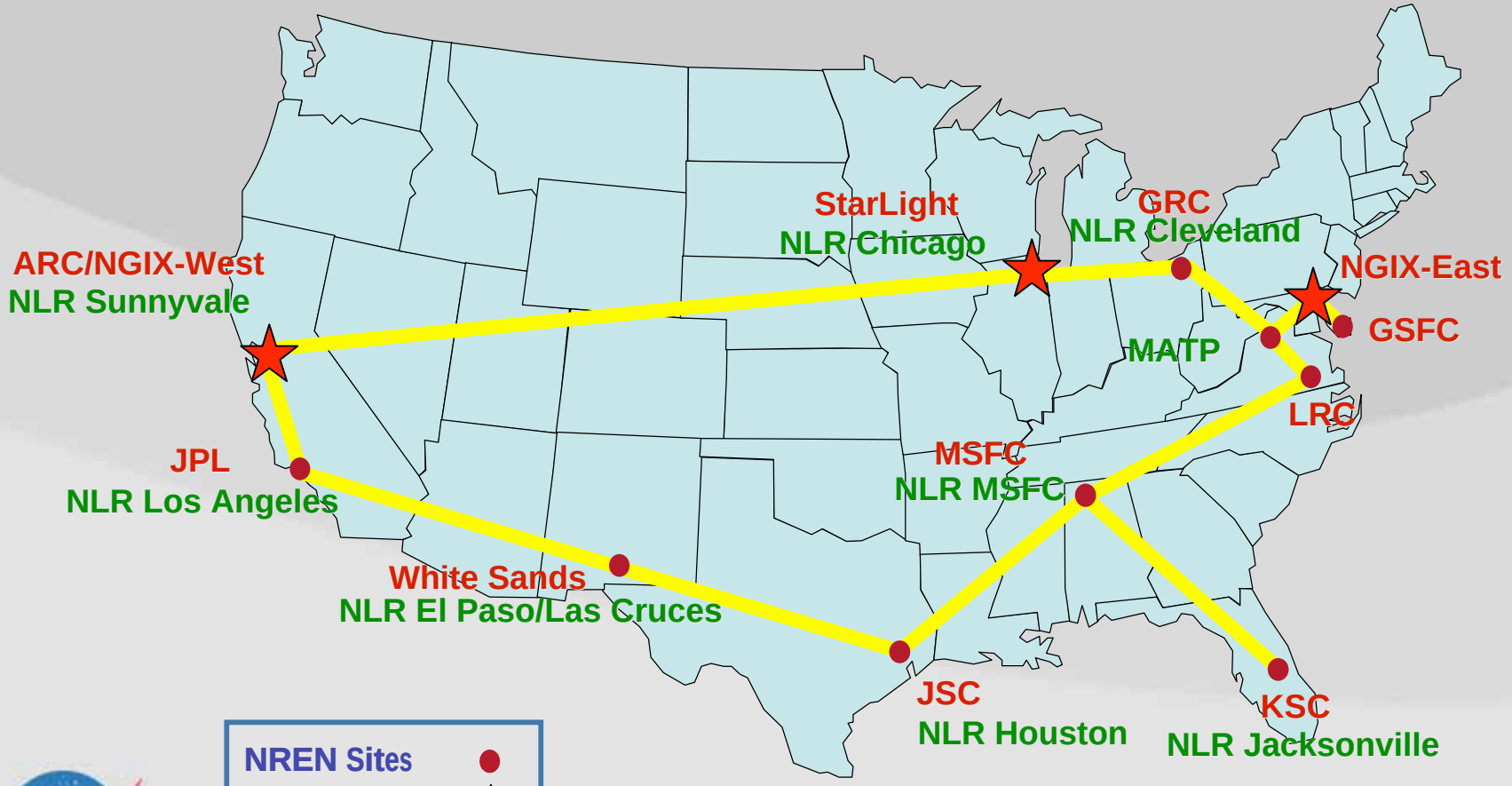
CCSP Reanalysis

- Leveraging REASoN CAN, Synergy, and DAAC resources
- Producing for CCSP Synthesis and Assessment report 1.3: “Re-analyses of historical climate data for key atmospheric features. Implications for attribution of causes of observed change”
- Requiring ordering large volume of data from the DAACs, reformatting, mining and thinning the data, and delivering to the compute platform
- Requiring a significant amount of compute cycles
- Requiring dissemination of large volume of data

Network Implementation Plan

	Oct 04 Phase 0 1 Gbps	Jul 05 Phase 1 10 Gbps	Jan 06 Phase 2 10 Gbps	Jul 06 Phase 3 10 Gbps
ARC	x	x		
JPL	x	x		
GSFC		x		
Starlight		x		
MSFC			x	
LRC			x	
GRC			x	
JSC			x	
NGIX-West		x		
NGIX-East		x		
KSC				x
WSTF				x

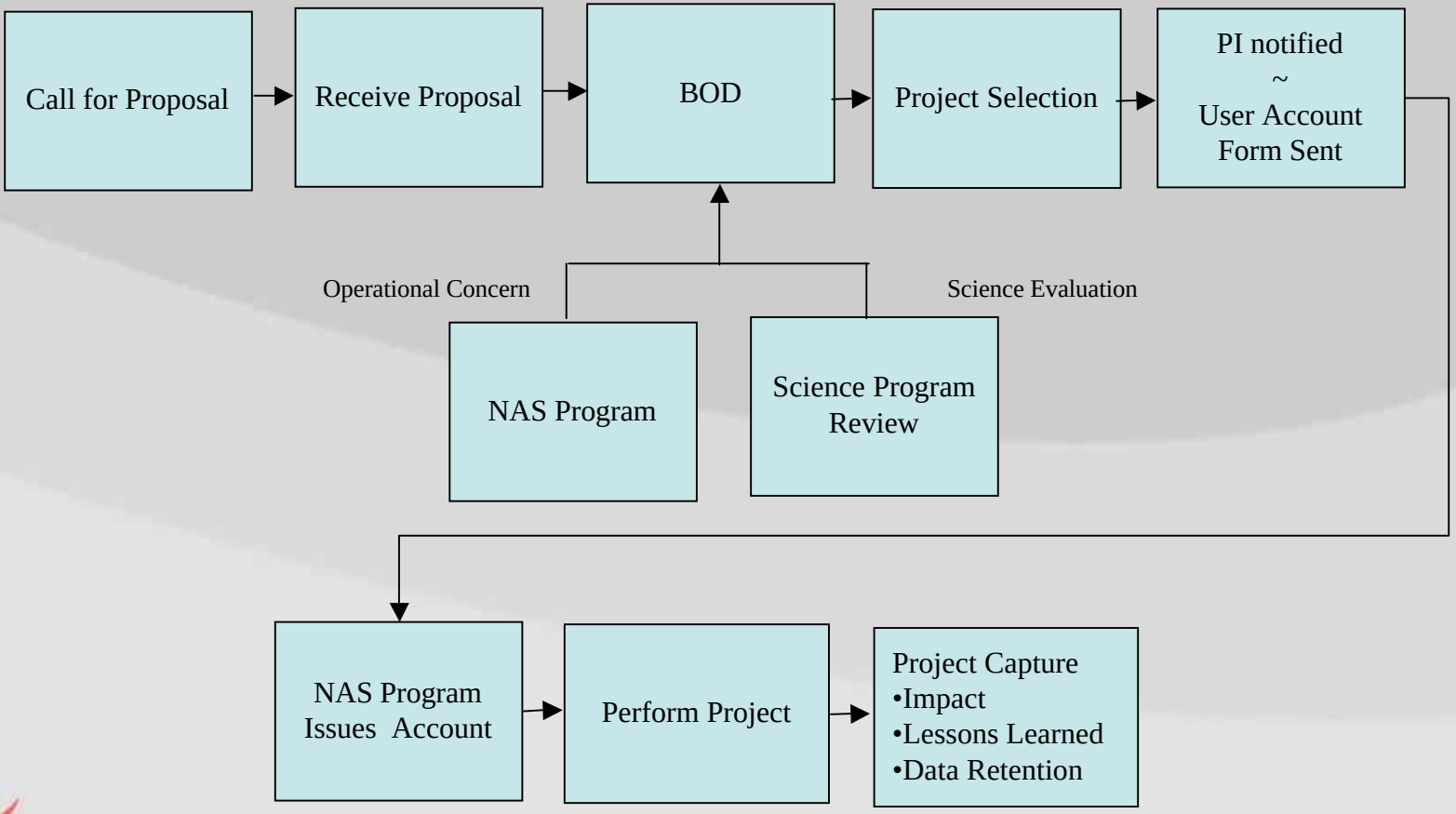
Phase 3 – July 2006



Project Columbia Governance

- NASA Deputy Administrator memo to mission directorates to form a Project Columbia “Board of Directors” organized by CIO
- BOD has been established and chaired by the Deputy CIO
- Discussions on high level allocation, budget advocacy, and Interagency relationship/partnership
- Joe Bredekamp represents Science Mission Directorate on the BOD
- Science Mission Directorate (Joe Bredekamp) takes the leadership in developing a OneNASA call for Project Columbia proposals
- NASA will involve other agencies in the call for proposals

Planning: Call for Project Columbia Computing Proposals



Proposed Allocation Vehicles

1. Project Columbia Call for Proposals
   - The Project Columbia call for proposals will NOT be a NRA
   - Only Project Columbia computing resources will be allocated
   - No supplement resources ($ and FTE) for data analysis
   - Reviewed and concurred by the science programs and external experts
   - Planning to release call for proposals every 6 months
   - Allocations awarded for 12 months

2. Modeling, Analysis, and Prediction (MAP) NRA
   - For the Earth Science, the MAP NRA will allocate significant Earth Science computing resources after the anticipated selection in April, 2005
   - The MAP NRA called for greater integration of a community Earth system model and data assimilation system under the Earth System Modeling Framework
   - MAP NRA Appendix A specifically dealt with computing resource requests

Selection Priority

- Selection Priority:
  - NASA funded projects
    - Requiring the validation from the science programs
    - Clearing for operational concerns
  - Interagency projects
    - Requiring the validation from the science programs
    - Clearing for operational concerns
  - New projects
    - Require science and operation reviews
- Project Priority:
  - Projects that will couple (1) models, (2) NASA data, (3) network distributed resources, and (4) data products dissemination
  - Projects that contain more than one aspect of the above

Summary

- Project Columbia is an opportunity to organize a more coherent OneNASA strategy for high-end computing
- Science Mission Directorate is engaged in the active management of NASA's high-end computing assets for the benefit of science
- The governance model is a work in progress
- The first OneNASA call for proposals is being developed
- SMD is facing continuing challenge in sustaining computing