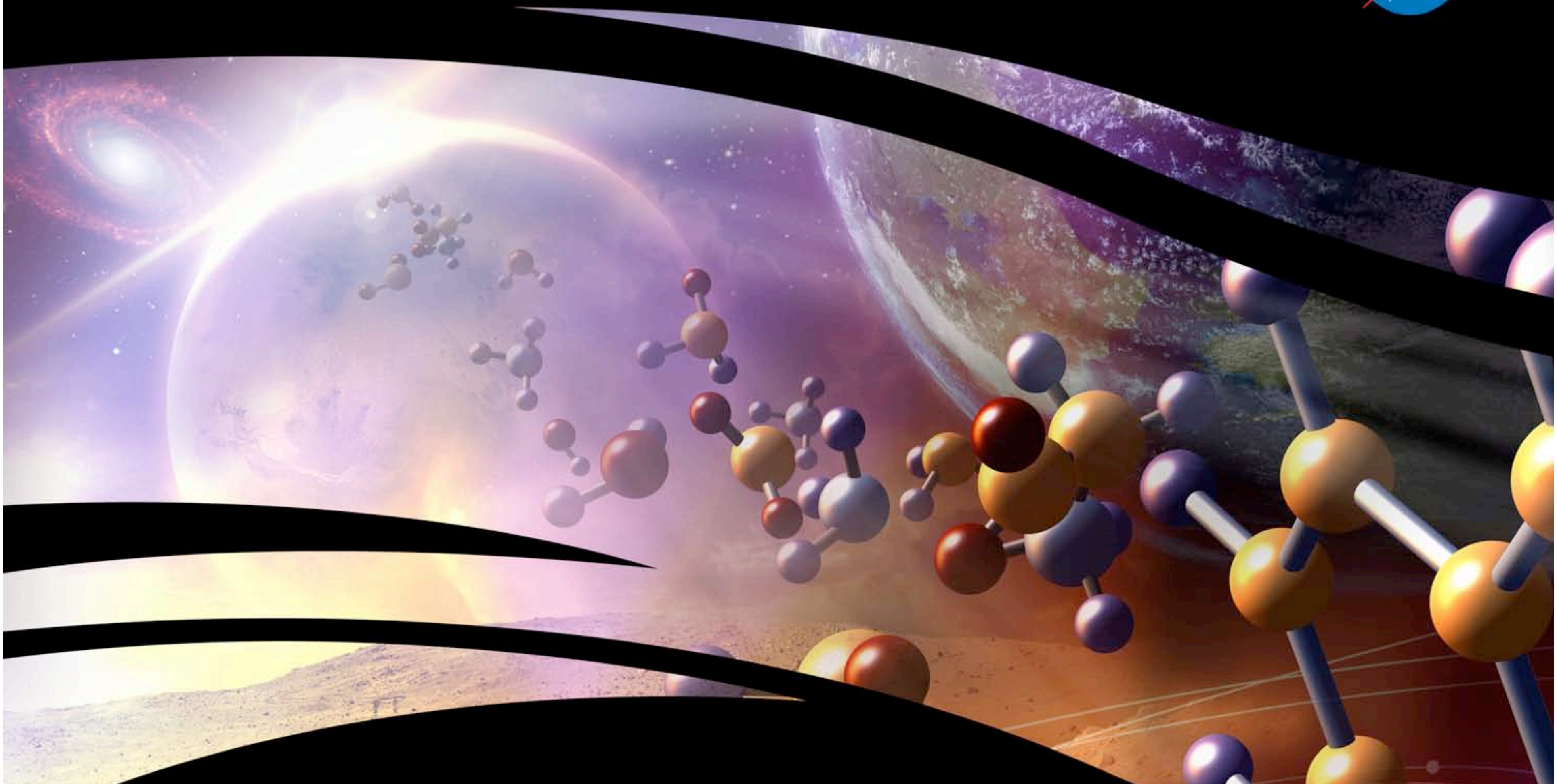
National Aeronautics and Space Administration



Charge to Workshop Participants

Tsengdar Lee
July 29, 2008

2002 Computational Technology Workshop Summary

NASA science requires major advances in computational technology

- NASA's unique driver is the data

Achievement of the prediction goals will require coordinated investments in science advancement and computational technologies

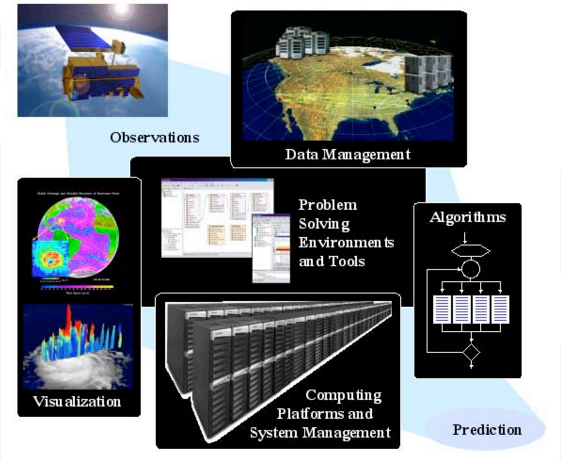
- Key stressing science applications are already identified
- Data management, application throughput, and problem solving environments are common across disciplines
- Science applications will build upon industry best practices, standards, and commercial offerings - But industry will not provide key technologies required to enable the stressing applications
- Coordinated investment in these identified technologies will benefit all the disciplines

Continued, focused investment in a science driven technology development program is required for success in the ESE

Report from the
Earth Science Enterprise

Computational Technology Requirements Workshop

April 30 - May 1, 2002



Weather, Climate, and Solid Earth panels defined capabilities needed to achieve NASA prediction goals in 2010

These capabilities were analyzed for stressing technology requirements

Technology Cross-Cut of Gaps Identified

A. Computing Platforms

B. Data Management

C. Programming Environment and Tools

D. Distributed Computing

E. Other Requirements

A. Computing Platform Throughput Required

	Stressing Model	Single Image Throughput	Estimated Capacity Required
Weather	10 Day Forecast Atmosphere:10 km horizontal, 100 levels vertical 10^{11} observations	20 Tflops	400 Tflops
Climate	S-I Prediction Atmosphere: 25 km horizontal Ocean: 6 km horizontal	5 Tflops	100s Tflops
Solid Earth	Earthquake Fault Slip 16M finite elements 100k boundary elements	2 Tflops	10s – 100 Tflops

Sustained Throughput and Capacity Requirements

- Vendors are expected to offer 20-50 Gflops processors, platforms with 10,000 processors, plenty of memory and storage
- Gaps are in achievable applications performance
- Needs:
  - Single processor application performance at a significant fraction of peak
  - Application scalability to thousands of processors
  - I/O performance that scales with the application performance

B. Data Management Requirements

- Data volume is expected to be overwhelming and heterogeneous in format
- Model output data management is the problem

	Observational Data	Access Modes Rates	Output Data	Storage Term/Re-access Mode
Weather Forecast	**1 TB/day Multiple Sources Continuous**	**Streamed input 20 GB/s**	**10 PB/day – Archival 10 TB/day – external distribution**	**Medium – Long Catalogued**
Climate Modeling	**10s of GB from archival sources**	**Data archive request 2 GB/s (latency tolerant)**	**100s TB/day**	**50% Short term - Immediate analysis 50% Medium term - Catalogued**
Solid Earth Research	**100s of GB/day Distributed sources**	**Distributed archives – low latency access**	**1 PB/day – ingested into distributed archives**	**Medium – Long Catalogued access**

- Vendors expected to provide physical storage solutions
- Gaps are in management, distribution of data volume
- Needs:
  - Uniform, location independent service for identifying, managing, and accessing metadata and raw data
  - Data transport performance that scales to consumer requirements
  - Low latency random access

C. Programming/Problem Solving Environment Requirements

- Applications to become much more complex - No single person will understand all of the details
- New applications need to be implemented in a month instead of a year
- Performance (efficiency) must be maintained without heroic efforts
- Ensemble executions, distributed application executions must be transparently manageable

- *No discipline specific vendor offerings expected in this area*
- Needs:
  - Application frameworks/composable component architectures
  - Platform independent program design and execution environment
  - Highly efficient applications that scale to 1,000s of processors without heroic effort

D. Distributed Computing Requirement

- Integration of geographically distributed data servers, computing assets, and users will be the norm
- Assets need to be unified in a seamless environment for maximum productivity
- Transparent, reliable data transport layer for interservice communication is required
- *Unpredictable vendor offerings in this area in 2010*
- Currently, there is a multi-agency investment (NSF, DOE, NASA) in this area (GRID computing)
- Needs:
  - Uniform, seamless, transparent access and programming environment

E. Other Requirements

- Scalable, efficient, reusable implementations of common algorithms
- Real time visualization of Terabytes of data
- Scalable data mining applications
- Computing platform systems management

Charge to the Workshop Participants

In the context of NASA data driven scientific research and analysis:

- What are the Key stressing science drivers?
- What are the computing cycle, storage, and networking needs?
- What are the data management, application throughput, and problem solving environments needs?
- Is there any specialized human capital investment required?
- Science applications will build upon industry best practices, standards, and commercial offerings - But industry will not provide key technologies required to enable the stressing applications. What do you think we can do to mitigate the risks?
- How do we best coordinate investment in these identified technologies to benefit all the disciplines?
- Other recommendations?