This subtopic seeks advances in tools, techniques, and technologies for distributed information systems and large-scale numerical simulation. The goal of this work is to create an autonomous information and computing environment that enables NASA scientists to work naturally with distributed teams and resources to dramatically reduce total time-to-solution (i.e., time to discovery, understanding, or prediction), vastly increase the feasible scale and complexity of analysis and data assimilation, and greatly accelerate model advancement cycles. Areas of interest follow below.

Distributed Information Systems

- Core services (autonomous software systems) for automated, scalable, and reliable management of distributed, dynamic, and heterogeneous computing, data, and instrument resources. Services of interest include those for authentication and security, resource and service discovery, resource scheduling, event monitoring, uniform access to compute and data resources, and efficient and reliable data transfer;
- Services for management of distributed, heterogeneous information, including replica management, intuitive interfaces, and instantiation on demand or "virtualized data." These services would be used, for example, to access and manipulate NASA's wealth of geospatial and remote sensing data;
- Science portals for cross-disciplinary discovery, understanding, and prediction, encapsulating services for single sign-on access, semantic resource and service discovery, workflow composition and management, remote collaboration, and results analysis and visualization; and
- Tools for rapidly porting and hosting science applications in a distributed environment. These applications should be written for an integrated, or workstation, environment using standard programming languages or tools such as Matlab, Interactive Data Language (IDL), or Mathematica.

Large-Scale Numerical Simulation

- Tools for automating large-scale modeling, simulation, and analysis, including those for managing computational ensembles, performing model-optimization studies, interactive computational steering, and
maintaining progress in long-running computations in spite of unreliable computing, data, and network resources;

- Tools for computer system performance modeling, prediction, and optimization for real applications;
- Techniques and tools for application parallelization and performance analysis;
- Tools for effective load balancing, and high reliability, availability, and serviceability (RAS) in commodity clusters and other large-scale computing systems; and
- Novel supercomputing approaches using FPGAs, graphics processors, and other novel architectures and technologies.