Large-scale data-intensive physics computing

Amber Boehnlein, Stanford University

All experimental and observational science areas are seeing dramatic increases in data volumes and a need for sophisticated data management, analysis and visualization techniques. There is increasing interplay of simulation and experiment observation that requires the ability to treat simulated and experimentally collected data on equal footing. The scale of the computing requirements and the distributed nature of the collaborations in high energy and nuclear physics experiments has led to the development of highly organized computing models. The successful deployment of this computing paradigm was a major factor in the ability of the Large Hadron Collider collaborations to rapidly achieve key physics goals, such as the discovery of the Higgs Boson. The Large Synoptic Space Telescope is an example of a new instrument that will push the boundaries of research data.

In contrast, the data rates and computing power traditionally required by experiments mounted at conventional light source end stations have been relatively modest and adequately addressed within the individual experimental groups. Due to advances in detector technology, the use of computer simulations to design experiments and a desire for near real-time feedback during data collection, light source users are experiencing significant increases in data rates and computational needs. This trend, coupled with the development of open data policies, is leading to a need for more formal computing paradigms.