TeraGrid Usage Modalities

The TeraGrid represents one of an emerging class of entities that can be referred to as "production cyberinfrastructures." These cyberinfrastructures, which also include the Open Science Grid in the US, DEISA and EGI in Europe, and RENKEI in Japan, harness distributed resources, including high-performance computing (HPC), storage, and specialized computing systems. These combinations of integrated resources make possible a variety of user and usage scenarios. To better support our user community, TeraGrid wants to understand how users are actually using these resources to achieve their scientific objectives, so that we can make changes in the TeraGrid environment to improve operations and services. Because TeraGrid will be transitioning into a new project (eXtreme Digital, or XD) in mid-2011, much of this data gathering will be given to the new project and can be used in planning its growth. A great deal of attention has gone into the study of HPC workloads, with the goals of improving scheduler performance and maximizing resource utilization. A previous study of the TeraGrid's workload patterns shows that, in many ways, the TeraGrid's HPC resources demonstrate many of these same patterns, both on individual systems and across the federation. However, while standard workload analyses can describe what users are doing on TeraGrid's HPC systems, they cannot easily be used to understand why users do what they do and how they leverage multiple types and instances of cyberinfrastructure (CI) resources. We have labeled these generalized, CI-spanning classes of user activity "usage modalities" and have begun efforts to instrument TeraGrid infrastructure services so that we can conduct quantitative analysis of user behavior in terms of these modalities. The paper is organized as follows. We first define the modalities we think are important to measure. We then discuss what is possible to measure in general, and on the TeraGrid. We next explain the characteristics of the modalities we have selected: user intent, when-to-run, submission mechanism, resources, job coupling, support, and level of software development. We finally discuss what else we would like to do, both internally on TeraGrid and with other infrastructures, and then conclude.

Primary authors : KATZ, Daniel S. (University of Chicago & Argonne National Laboratory)

Co-authors : HART, David (NCAR) ; JORDAN, Chris (TACC) ; MAJUMDAR, Amit (SDSC/UCSD) ; NAVARRO, Jp (University of Chicago & Argonne National Laboratory) ; SMITH, Warren (TACC) ; TOWNS, John (NCSA) ; WELCH, Von (Von Welch Consulting, LLC) ; WILKINS-DIEHR, Nancy (SDSC/UCSD)