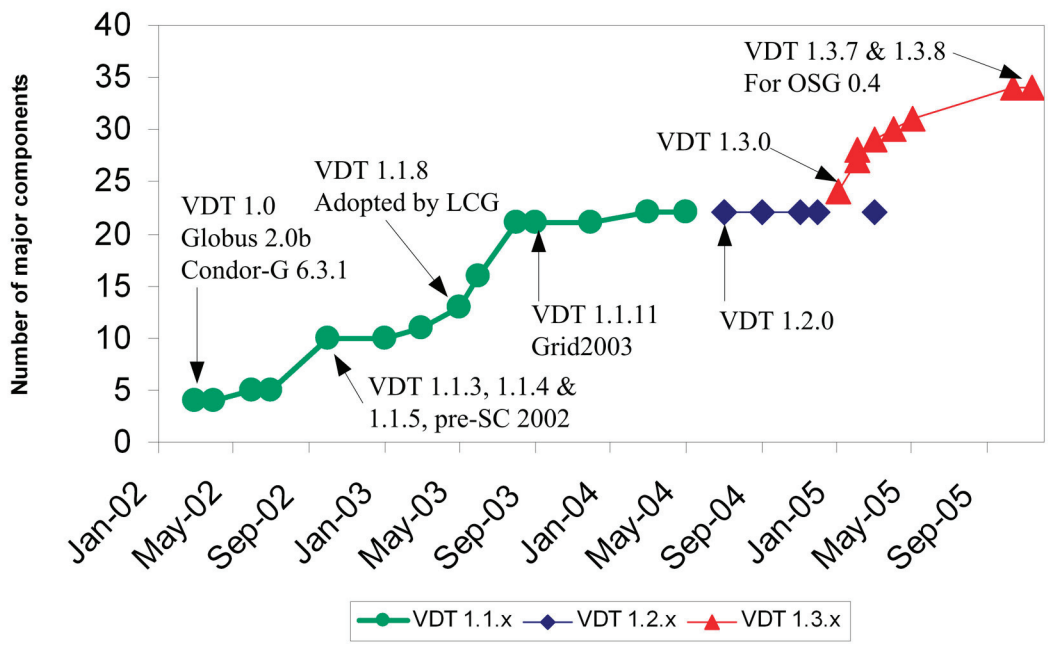
Computer Science on the Open Science Grid

Ongoing computer science research is essential to achieving the mission and vision of the OSG. At the same time, the OSG provides an unparalleled opportunity for computer scientists to test new concepts and advances in a live, demanding computational science laboratory. Here are a few of the research programs actively engaged with the OSG.

Virtual Data Toolkit

The VDT is a collection of grid middleware that can be easily installed and configured. Its goal is to enable researchers to use or provide grid resources with a simple "shrink-wrapped" software installation. Originally created to serve as a delivery channel for grid technologies developed and hardened by the NSF-funded GriPhyN and iVDGL projects, the VDT now supports the OSG, the LHC Computing Grid Project (LCG), Enabling Grids for E-sciencE (EGEE) and several others. The VDT contains basic grid services such as Condor-G and Globus Toolkit 4, monitoring software such as MonALISA and the Generic Information Provider, security software such as VOMS, and a variety of other utilities. It is built using the NSF Middleware Initiative's Build and Test facility and is tested nightly.



Growth in number of VDT components over time. (Alain Roy)

Clarens

The Clarens Grid-Enabled Web Services Framework is an open source, secure, high-performance portal for access to data and computational resources provided by computing grids. Clarens was developed as part of a wide-area network Grid-enabled Analysis Environment for collaborative analysis of data generated by the CMS detector at CERN. The server-side application environment provides a standard, easy-to-use framework for developing grid services, with mechanisms for PKI-based authentication and authorization based on access control lists in the context of distributed VOs. Services are automatically published for discovery using a global service registry, and can be used from a variety of clients, including Web browsers, scripts and full-scale applications.

DOE Office of Science

sc05-inserts.indd 11 10/27/2005 7:06:43 PM

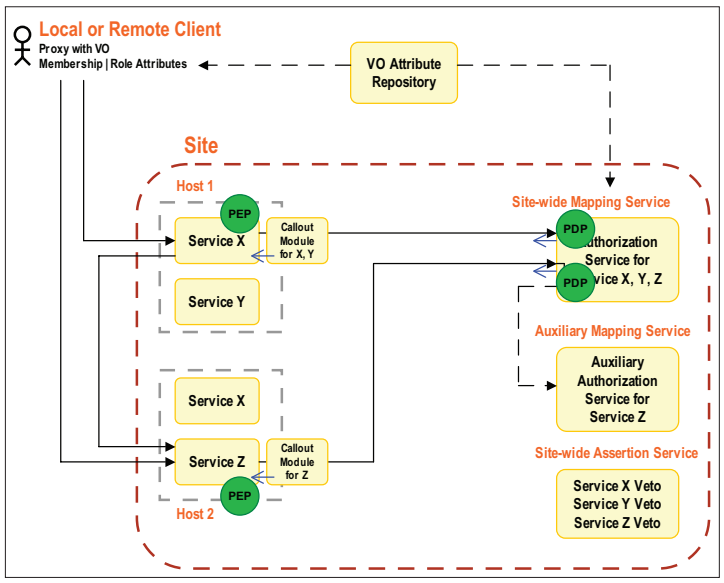
Computer Science on the Open Science Grid

Virtual Cluster and Workspaces

Two of the most important challenges facing the OSG are dynamically providing an execution environment required by a specific application on an arbitrary resource, and enforcing limits on the resources this application may consume. OSG computer science researchers are investigating two solutions to these problems using virtualization: an Edge Services Framework (ESF) allowing authorized clients to deploy a gateway to the site processing requests from remote grid clients; and a service to deploy a cluster of virtual machines as a workspace, configured to support the needs of OSG applications. Such clusters can be deployed on demand within seconds, and be allocated a lifetime based on need. ESF uses the Globus Toolkit's Workspace Service to enable a VO administrator to dynamically deploy virtual machines representing infrastructure nodes for different communities.

GriPhyN Virtual Data System

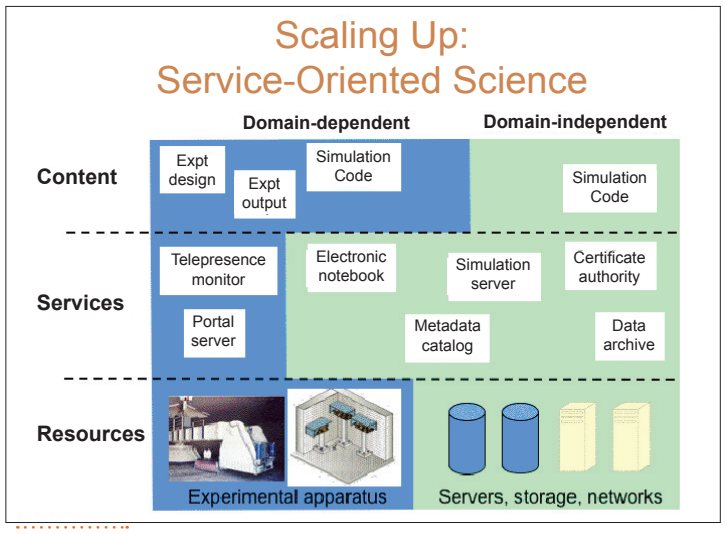
The VDS provides a set of tools for expressing, executing and tracking the results of scientific workflows. Workflows consist of graphs of application (and soon, of service) invocations, and can be expressed in a location-independent, high-level Virtual Data Language which frees the workflow from specifying details of the location of files and programs in a distributed environment. VDL workflows can be executed in a variety of environments ranging from the desktop to the OSG or the TeraGrid.



(Abhishek Rana and Frank Wuerthwein)

Core Platform Research

Research in the core technologies of Condor and Globus continues. Condor research focuses on integrating data transport into scientific workflows, and Globus research addresses problems in advanced security models, end-to-end data transport solutions and service-oriented architectures.



(Ian Foster)

DOE Office of Science

sc05-inserts.indd 12 10/27/2005 7:06:45 PM