

Open Science Grid

An Introduction

The Open Science Grid (OSG) enables collaborative science by providing a national cyber-infrastructure of distributed computing and storage resources. The goal of the OSG is to transform processing and data intensive science through a cross-domain, self-managed, nationally distributed cyber-infrastructure that brings together campus and community resources. This system is designed to meet the needs of Virtual Organizations (VOs) of scientists at all scales. OSG is jointly funded by the Department of Energy and the National Science Foundation to build, operate, maintain, and evolve a facility that will meet the current and future needs of large scale scientific computing. To meet these goals, OSG provides common services and support, a software platform, and a set of operational principles that organizes users and resources into Virtual Organizations.

Virtual Organizations

Virtual Organizations (VOs) are at the heart of OSG principles and its model for operation. VOs are a collection of researchers who join together to accomplish their goals; typically they share the same mission, but that is not a requirement for establishing an OSG VO. A VO joins OSG to share their resources, computing and storage, with the other OSG VOs and to be able to access the resources provided by other OSG VOs as well as share data and resources with international computer grids (e.g. EGEE). The resources owned by a VO are often geographically distributed; a set of co-located resources are referred to as a site and thus a VO may own a number of sites. There are two key aspects of VOs: 1) the user community within a VO that submits jobs into the OSG; and 2) the set of computing and storage resources that are owned by a VO and connected to the OSG. In some cases, VOs do not bring resources to OSG and are only users of available resources on OSG.

A key principle in OSG is the autonomy of VOs that allows them to develop an operational model that best meets their science needs; this autonomy applies both to their user community and sites. OSG requires each VO to establish certain roles (i.e. VO manager, VO admin, VO Security Contact) and agree to a set of policies (e.g. Acceptable Use Policy) which allow operation of the OSG as a secure and efficient grid. VOs administer, manage, and support their own user community. In addition, many VOs provide common software infrastructure designed to meet the specific needs of their users. VOs as providers of resources also have great autonomy in building and

operating their sites. Sites use the OSG software stack to provide the “middleware layers” that make their sites ready for connection to the OSG. Sites set policies on how their resources will be used by their own users and other VOs; the only requirement is that sites support at least one other VO but the site controls the conditions under which that resource is available. OSG does not tightly restrict what hardware or operating system software a VO may supply or what software it may use to access OSG or provide resources on OSG: they are autonomous and are allowed to make such choices as long as they meet the basic requirements. This autonomy allows a VO to build its computing resource to meet its specific needs and makes it more likely that a VO will choose to join OSG because it doesn’t have to compromise its own needs to do so.

Software Platform

The primary goal of the OSG software effort is to build, integrate, test, distribute, and support a set of common software for OSG administrators and users. OSG strives to provide a software stack that is easy to install and configure even though it depends on a large variety of complex software.

The key to making the OSG infrastructure work is a common package of software provided and supported by OSG called the OSG Virtual Data Toolkit (VDT). The VDT includes Condor and Globus technologies with additional modules for security, storage and data management, workflow and other higher level services, as well as administrative software for testing, accounting and monitoring. The needs of the domain and computer scientists, together with the needs of the administrators of the resources, services and VOs, drive the contents and schedule of releases of the VDT. The OSG middleware allows the VOs to build an operational environment that is customized to their needs.

The OSG supports a heterogeneous set of operating systems and versions and provides software that publishes what is available on each resource. This allows the users and/or applications to dispatch work to those resources that are able to execute it. Also, through installation of the VDT, users and administrators operate in a well-defined environment and set of available services.

Common Services and Support

To enable the work of the VOs, the OSG provides direct staff support and operates a set of services. These functions are available to all VOs in OSG and provide a foundation for the specific environments built, operated, and supported by each VO; these include:

1. Information, accounting, and monitoring services that are required by the VOs; and forwarding of this information to external stakeholders on behalf of certain VOs,
2. Reliability and availability monitoring used by the experiments to determine the availability of sites and to monitor overall quality,
3. Security monitoring, incident response, notification and mitigation,
4. Operational support including centralized ticket handling,
5. Collaboration with network projects (e.g. ESNET and Internet2) for the integration and monitoring of the underlying network fabric which is essential to the movement of peta-scale data,
6. Site coordination and technical support for VOs to assure effective utilization of grid connected resources,
7. End-to-end support for simulation, production, analysis and focused data challenges to enable the science communities accomplish their goals.

These centralized functions build centers of excellence that provide expert support for the VOs while leveraging the cost efficiencies of shared common functions.

OSG Today (April 2009)

OSG provides an infrastructure that supports a broad scope of scientific research activities, including the major physics collaborations, nanoscience, biological sciences, applied mathematics, engineering, and computer science. OSG does not own any computing or storage resources, but instead they are all contributed by the members of the OSG Consortium and are used both by the owning VO and other VOs; recent trends show that about 20-30% of the resources are used on an opportunistic basis by VOs that do not own them.

With about 80 sites and 30 VOs, the usage of OSG continues to grow; the usage varies depending on the needs of the stakeholders. During stable normal operations, OSG is providing over 600,000 CPU wall clock hours a day with peaks occasionally exceeding 800,000 CPU wall clock hours a day; approximately 100,000 to 200,000 opportunistic wall clock hours are available on a daily basis for resource sharing.