# What are good computations for the OSG resource?

OSG supports XSEDE users by providing a Virtual Cluster that forms an abstraction layer to access the distributed OSG infrastructure. Jobs submitted into the OSG Virtual Cluster will be executed on machines at several remote physical clusters. These machines may differ in terms of computing environment from the submit node. Therefore it is important that the jobs are as self-contained as possible by generic binaries and data which can be either carried with the job, or staged on demand.

Many high throughput computing workloads work well on OSG: parameter sweeps, Monte Carlo simulations, and bag-of-tasks. Note that OSG handles mostly serial applications.

**Please consider the following guidelines for what good workloads would look like:**

* Model should preferably be single threaded, using less than 2 GB memory and each invocation should run for 4-12 hours.
* Compute sites in the OSG can be configured to use pre-emption, which means jobs can be automatically killed if higher priority jobs enter the system. Pre-empted jobs will restart on another site, but it is important that the jobs can handle multiple restarts.
* Binaries should preferably be statically linked. However, dynamically linked binaries with standard library dependencies, built for a 64-bit Red Hat Enterprise Linux (RHEL) 5 machines will also work. Also, interpreted languages such as Python or Perl will work as long as there are no special module requirements.
* Input and output data for each job should be < 10 GB to allow them to be pulled in by the jobs, processed and pushed back to the submit node. Note that the OSG Virtual Cluster does not currently have a global shared file system, so jobs with such dependencies will not work.

**The following are examples of computations which are not good matches for OSG:**

* Tightly coupled computations, for example MPI based communication, will not work on OSG due to the distributed nature of the infrastructure.
* Computations requiring a shared file system will not work as there is no shared file system between the different clusters on OSG.
* Computations requiring complex software deployments are not a good fit. There is limited support for distributing software to the compute clusters, but for complex software, or licensed software, deployment can be a major task.