Supporting Shared Resource Usage for a Diverse User Community: the OSG experience and lessons learned

Marko Slyz, Gabriele Garzoglio, Tanya Levshina, Chander Sehgal, Mats Rynge
{slyz, garzoglio, tlevshin, cssehgal}@fnal.gov , rynge@isi.edu

Open Science Grid

Open Science Grid

Introduction

The Open Science Grid (OSG) is a consortium of more than 100 institutions including universities, national laboratories, and computing centers. OSG fosters scientific research and knowledge supporting the computational activities of more than 80 communities. Because of the federation of resources, during a cycle of low resource utilization at a particular institution other communities have the ability to use those idle resources "opportunistically" via a common interface. [1] To help enable this the OSG User Support group works with communities to port their computing operations to OSG.

OSG primarily supports running many simultaneous jobs that don't need low-latency communication to other jobs. This paradigm is called Distributed High Throughput Computing (DHTC).

Some problems can't easily be made to fit this paradigm, and may run better on supercomputers.

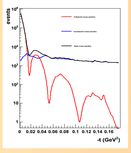
Other problems are more naturally suitable for DHTC. This poster discusses the computational requirements for some of these and a few workflows to run them on OSG. These workflows might serve as patterns for porting applications with similar requirements.

EIC

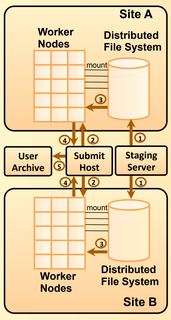
Science Goals

The Electron Ion Collider (EIC) is a proposed facility for studying the structure of nuclei. Engineers need a large amount of computations to define its design.[3]

Tobias Toll, Thomas Ullrich, BNL



Workflow

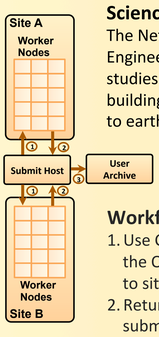


1. (Only once) Pre-stage a 1GB read-only file to each site. This way, that file does not need to be repeatedly transferred over the wide area network.
2. Submit the jobs to the sites with the pre-staged files. Transfer the application.
3. Jobs run and read the pre-staged files.
4. Condor transfers output data back to submit host.
5. User takes possession of the results.

Features

A lot of internal site transfers, as compared to the data transferred over the WAN.

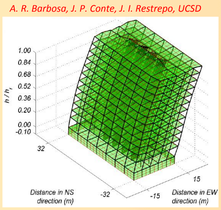
NEES



Science Goals

The Network for Earthquake Engineering Simulation (NEES) studies the response of buildings and other structures to earthquakes. [4]

A. R. Barbosa, J. P. Conte, J. I. Restrepo, UCSD



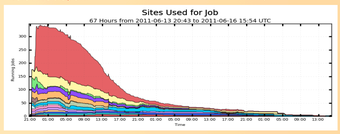
Workflow

1. Use Condor/glideinWMS to submit the OpenSees simulation application to sites. Condor transfers input data.
2. Return the data using Condor to the submit host.
3. Use Globus Online to transfer the data to the user's archive.

Features

Large amount of output (3GB / job). This caused multiple challenges, including jobs occupying computing resources after completion, waiting on the output data transfer queue.
This community would benefit from using the Pheno workflow, instead.

Derek Weitzel, UNL



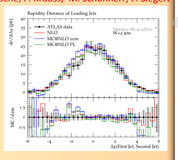
Completion time of 340 jobs at 16 sites. Jobs wait after completioon to transfer their output.

Pheno

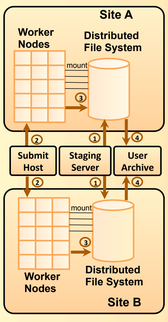
Science Goals

The Phenomenology group at SLAC ran an application called Sherpa that does multi-particle quantum chromodynamics calculations using Monte Carlo methods.

S. Höche, F. Krauss, M. Schönherr, F. Siegert



Workflow



1. (Only once) Stage software to each site.
2. Submit the jobs to the sites and transfer input data to local storage.
3. Transferring output data to local storage at site.
4. Users retrieves output data directly from the storage at each site.

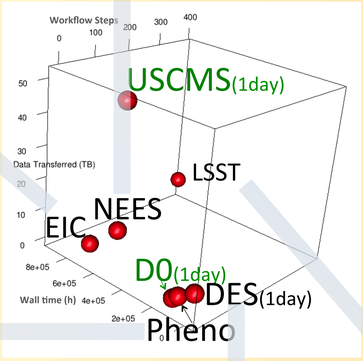
Features

Fair amount of output data. After the experience with NEES on large output, using a distributed data workflow.

Workflow Metrics

The following workflow metrics help illustrate the similarities between different communities:

- **Wall time**: Aggregate execution time of the computational campaign.
- **Data**: Aggregate transfer size of input, output and the executable.
- **Workflow Steps**: A workflow is a logical unit of computation, composed of one or more data processing steps. Each step runs as an independent Grid job.
- **Job Count**: the number of jobs in the campaign.



Project	Workflow Steps	Job Count	Wall time (h)	Data (TB)	Hours per job
Pheno at SLAC	1	9000	100000	1.9	0
EIC	1	178000	696000	3.2	5
LSST Simulation	380	380000	909000	5	7
NEES OpenSees	1	17000	509000	12	29
DES (1 day)	1	300	5000	5.4	16
US CMS (1 day)	10	102000	519000	50	5
D0 (non-local) (1 day)	1	18000	130000	1	7

Graph and table [2]:
The numbers in the "Data Transferred" column, and a few others, are estimates. Aim of the graph is to give a sense of the scale of workflows.
Data for US CMS and D0 on OSG are shown for comparison. The data is for only one day of their operations so that they're on the same scale as the OSG projects. The OSG projects, however, accumulate their time over the course of months.
Among the OSG sites, D0 runs preferentially on FermiGrid (250 TB/day transferred).

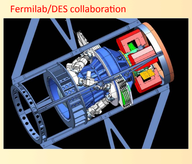
DES

Science Goals

To better understand the properties of dark energy, the Dark Energy Survey (DES) is planning to collect and process images of the night sky.

NASA/ESA via Wikipedia

Fermilab/DES collaboration



Fermilab

Workflow

For prototyping, we are using a workflow similar to the one for the Pheno group at SLAC. Final workflow will be complicated by large data requirements and end-to-end system complexity.
Work in collaboration with DES Data Management group. This project is being evaluated for running on OSG.

Features

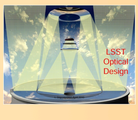
Large amount of data to transfer. Complex requirements and end-to-end Grid system.

LSST

Science Goals

Produce simulated images for the Large Synoptic Survey Telescope (LSST) project for use in verifying the LSST software. LSST will be an 8 meter wide-field telescope that will image the entire visible sky every few nights for 10 years.

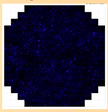
LSST Collaboration



Workflow

1. (Only once) Pre-stage star catalog and focal plane configuration files (15 GB) .
2. Submit 1 job to trim the pre-staged catalog file into 189 files, one per CCD chip in the camera.
3. Submit 2 x 189 jobs: simulate 1 image pair (same image with 2 exposures). Transfer "instance catalog" (telescope position, wind speed, etc.) with each job.
4. Gather output, perform bookkeeping, etc.

LSST Image Simulation team



Features

Needs input pre-processing → complex workflow. Large catalog file pre-staged at all sites. Workflow simplified at Purdue since then.

References

[1] "What is the Open Science Grid?", https://twiki.grid.iu.edu/bin/view/Documentation/WhatIsOSG
[2] The OSG Gratia Server, the PheEDEx monitoring system , and discussion with representatives from experiments.
[3] T. Toll, "Electron Ion Collider Simulations on OSG", https://twiki.grid.iu.edu/bin/view/Management/NovDec2011
[4] A. R. Barbosa, J. P. Conte, J. I. Restrepo, "Running OpenSees Production for NEES on OSG", https://twiki.grid.iu.edu/bin/view/Management/Oct2011Newsletter

Acknowledgments: Fermilab is operated by the Fermi Research Alliance, LLC under Contract No. DE-AC02-07CH11359 with the United States Department of Energy.
The Open Science Grid (http://www.opensciencegrid.org) is funded by the National Science Foundation and the Office of Science, Department of Energy (DOE).