Open Science Grid

OSG Resource Selection Service (ReSS)

Overview

- The ReSS Project (collaboration, architecture, …)
- ReSS Validation and Testing
- Project Status and Plan
- ReSS Deployment

Don Petravick for Gabriele Garzoglio
Computing Division, Fermilab

ISGC 2007

The ReSS Project

- The Resource Selection Service implements cluster-level Workload Management on OSG.
- The project started in Sep 2005
- Sponsors
  - DZero contribution to the PPDG Common Project
  - FNAL-CD
- Collaboration of the Sponsors with
  - OSG (TG-MIG, ITB, VDT, USCMS)
  - CEMon gLite Project (PD-INFN)
  - FermiGrid
  - Glue Schema Group

Open Science Grid

Motivations

- Implement a light-weight cluster selector for push-based job handling services
- Enable users to express requirements on the resources in the job description
- Enable users to refer to *abstract* characteristics of the resources in the job description
- Provide soft-registration for clusters
- Use the standard characterizations of the resources via the Glue Schema

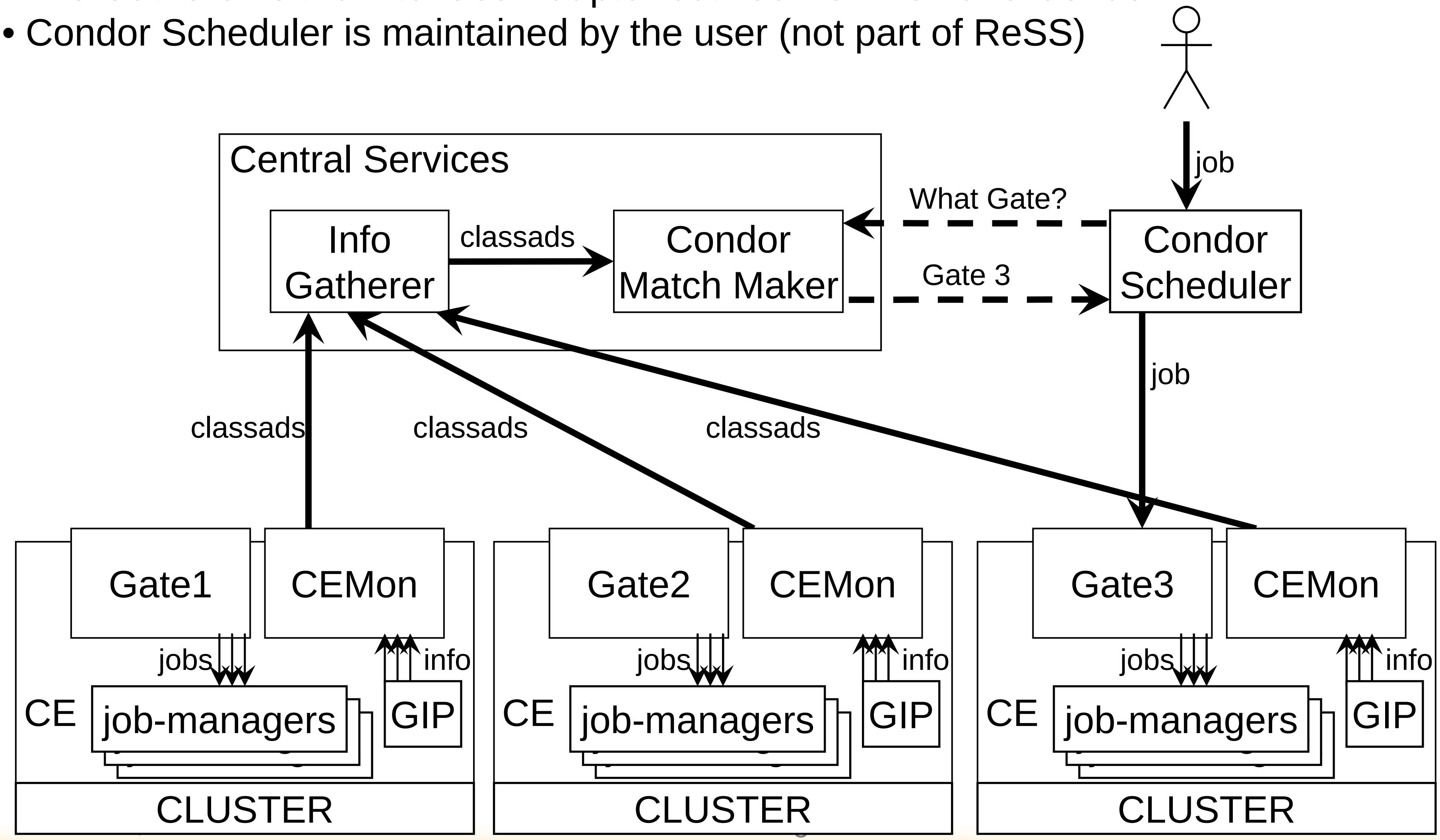
Technology

- ReSS basis its central services on the Condor Match-making service
  - Users of Condor-G naturally integrate their scheduler servers with ReSS
  - Condor information collector manages resource soft registration
- Resource characteristics is handled at sites by the gLite CE Monitor Service (CEMon)
  - CEmon registers with the central ReSS services at startup
  - Info is gathered by CEMon at sites running Generic Information Prividers (GIP)
  - GIP expresses resource information via the Glue Schema model
  - CEMon converts the information from GIP into old classad format. Other supported formats: XML, LDIF, new classad
  - CEMon publishes information using web services interfaces

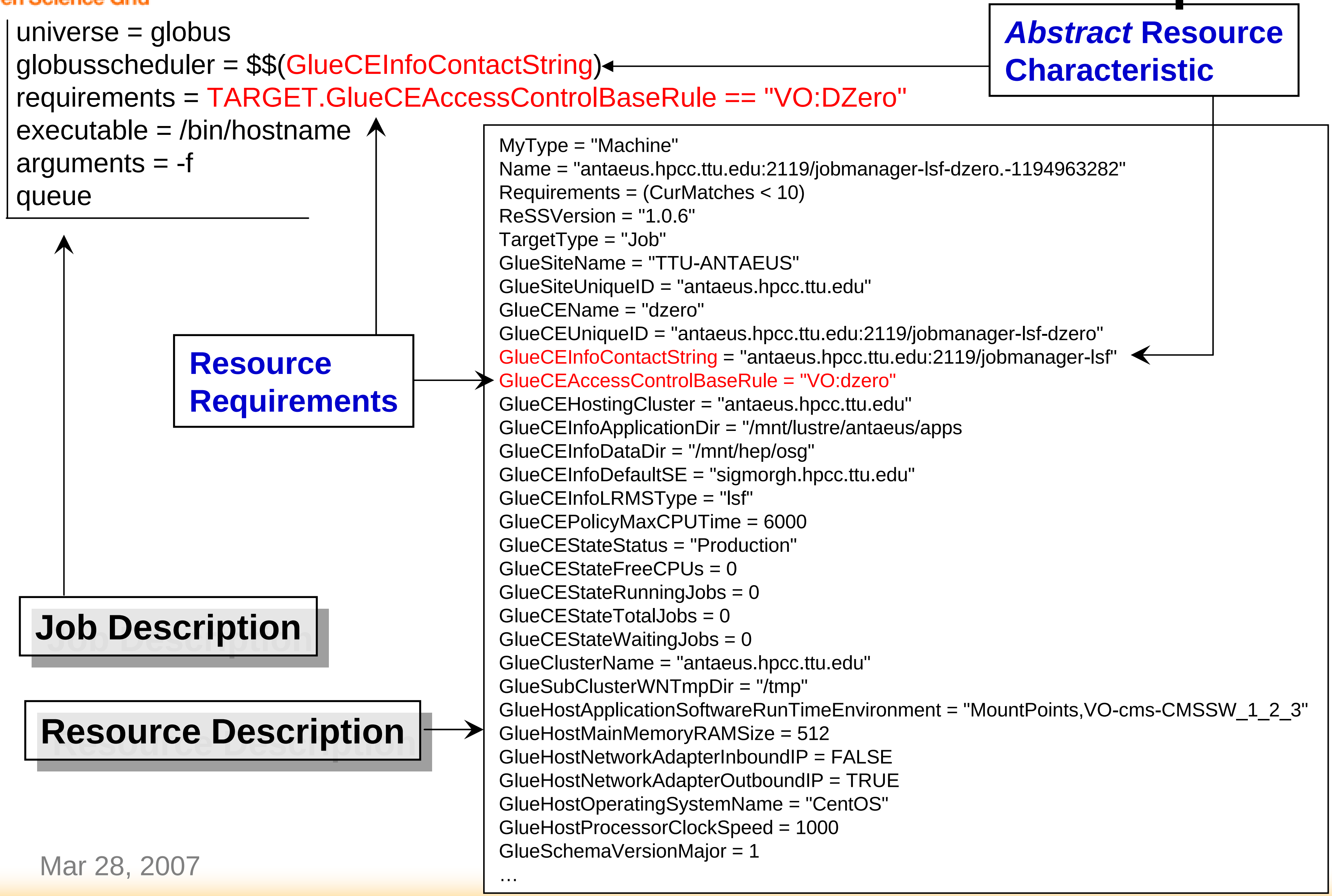
Open Science Grid

Architecture

- Info Gatherer is the Interface Adapter between CEMon and Condor
- Condor Scheduler is maintained by the user (not part of ReSS)



Resource Selection Example



```
universe = globus
globusscheduler = $$(GlueCEInfoContactString)
requirements = TARGET.GlueCEAccessControlBaseRule == "VO:DZero"
executable = /bin/hostname
arguments = -f
queue
```

***Abstract* Resource Characteristic**

Resource Requirements

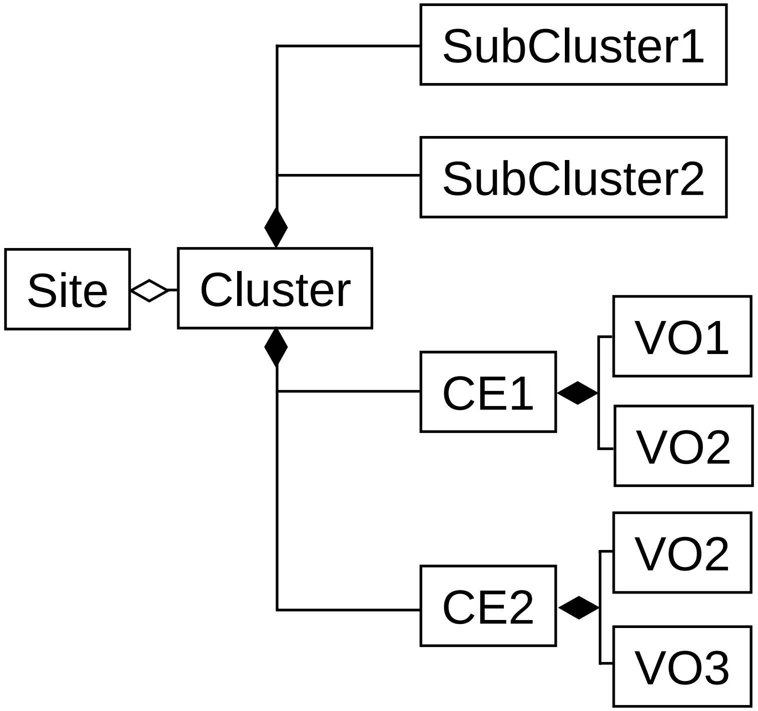
Job Description

Resource Description

```
MyType = "Machine"
Name = "antaeus.hpcc.ttu.edu:2119/jobmanager-lsf-dzero.-1194963282"
Requirements = (CurMatches < 10)
ReSSVersion = "1.0.6"
TargetType = "Job"
GlueSiteName = "TTU-ANTAEUS"
GlueSiteUniqueID = "antaeus.hpcc.ttu.edu"
GlueCEName = "dzero"
GlueCEUniqueID = "antaeus.hpcc.ttu.edu:2119/jobmanager-lsf-dzero"
GlueCEInfoContactString = "antaeus.hpcc.ttu.edu:2119/jobmanager-lsf"
GlueCEAccessControlBaseRule = "VO:dzero"
GlueCEHostingCluster = "antaeus.hpcc.ttu.edu"
GlueCEInfoApplicationDir = "/mnt/lustre/antaeus/apps
GlueCEInfoDataDir = "/mnt/hep/osg"
GlueCEInfoDefaultSE = "sigmorgh.hpcc.ttu.edu"
GlueCEInfoLRMSType = "lsf"
GlueCEPolicyMaxCPUTime = 6000
GlueCEStateStatus = "Production"
GlueCEStateFreeCPUs = 0
GlueCEStateRunningJobs = 0
GlueCEStateTotalJobs = 0
GlueCEStateWaitingJobs = 0
GlueClusterName = "antaeus.hpcc.ttu.edu"
GlueSubClusterWNTmpDir = "/tmp"
GlueHostApplicationSoftwareRunTimeEnvironment = "MountPoints,VO-cms-CMSSW_1_2_3"
GlueHostMainMemoryRAMSize = 512
GlueHostNetworkAdapterInboundIP = FALSE
GlueHostNetworkAdapterOutboundIP = TRUE
GlueHostOperatingSystemName = "CentOS"
GlueHostProcessorClockSpeed = 1000
GlueSchemaVersionMajor = 1
…
```

Open Science Grid

Glue Schema to old classad Mapping

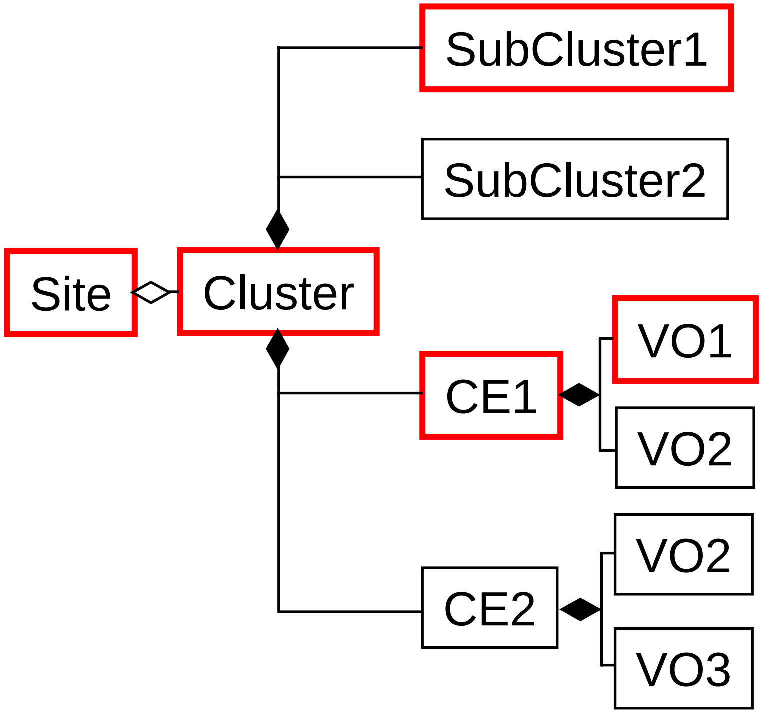


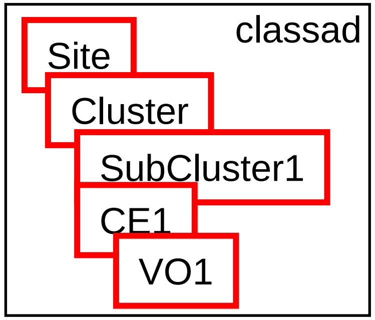
Mapping the Glue Schema “tree” into a set of “flat” classads:
all possible combination of (Cluster, Subcluster, CE, VO)

…

Open Science Grid

Glue Schema to old classad Mapping

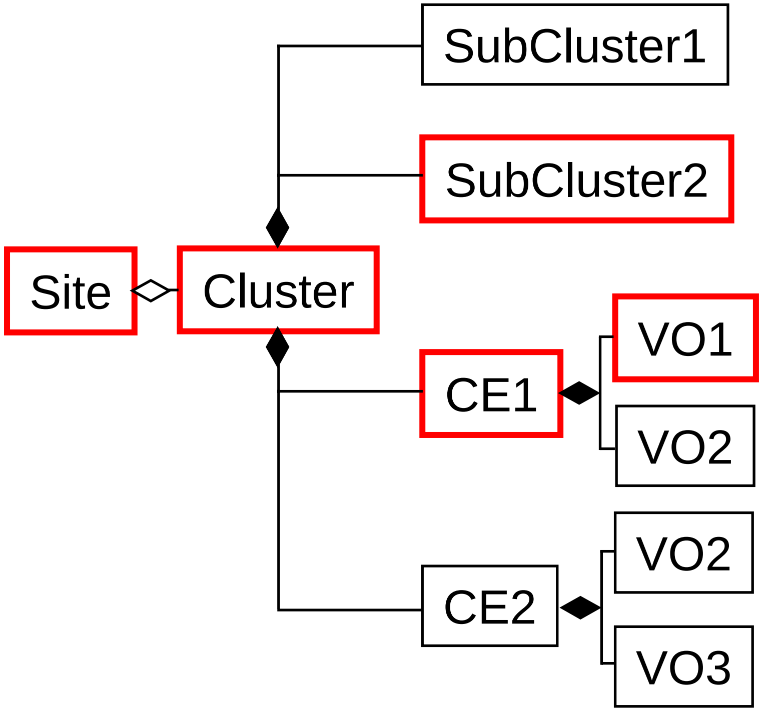


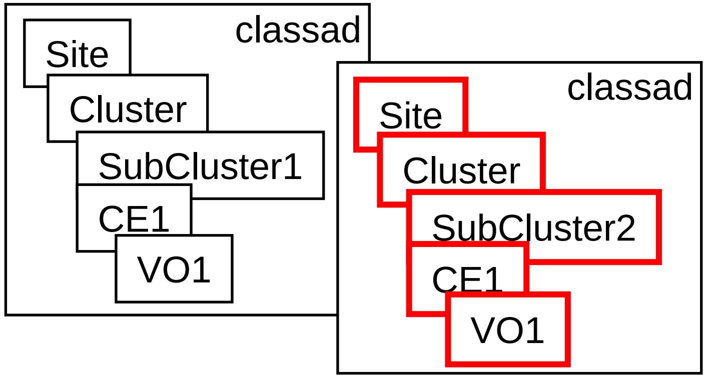


Mapping the Glue Schema “tree” into a set of “flat” classads:
all possible combination of (Cluster, Subcluster, CE, VO)

…

Glue Schema to old classad Mapping

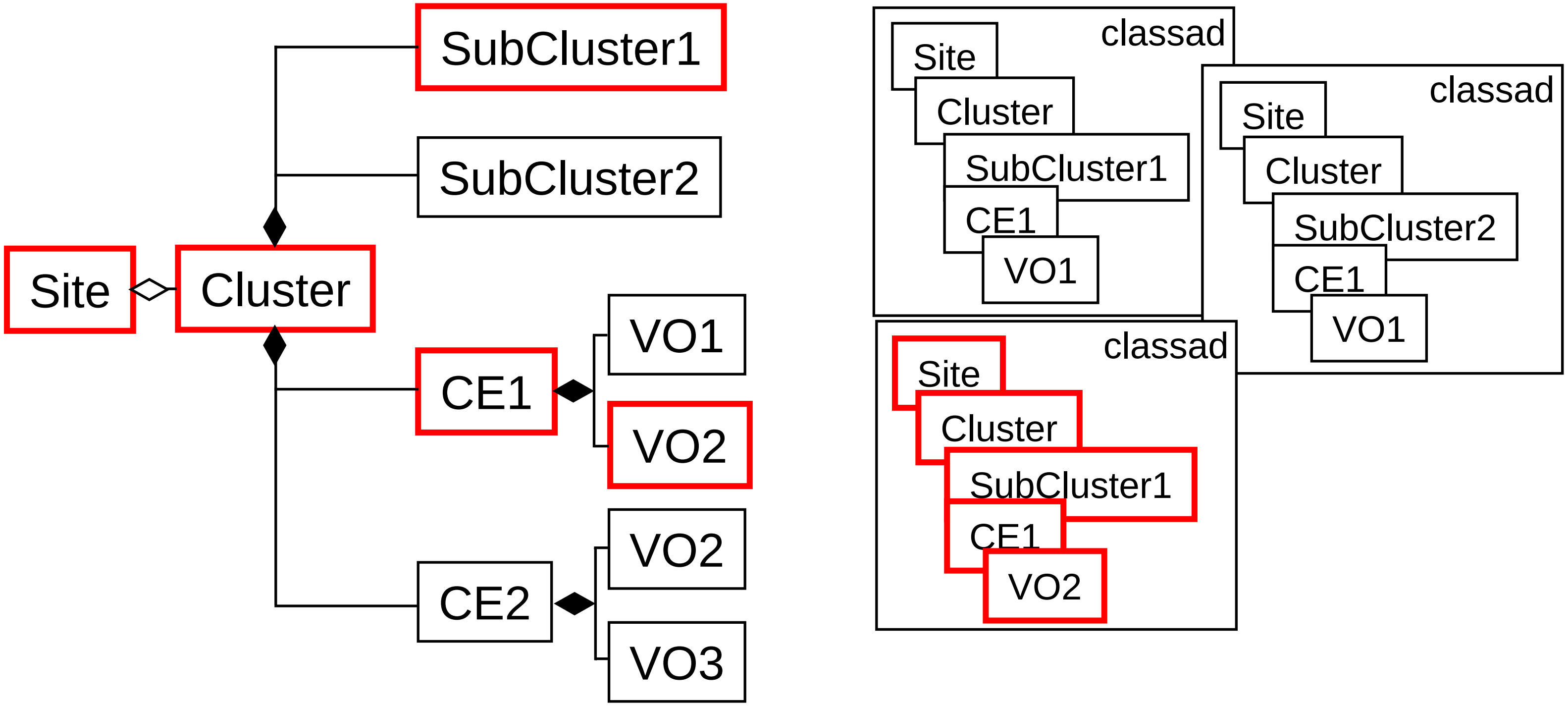




Mapping the Glue Schema “tree” into a set of “flat” classads:
All possible combination of (Cluster, Subcluster, CE, VO)

…

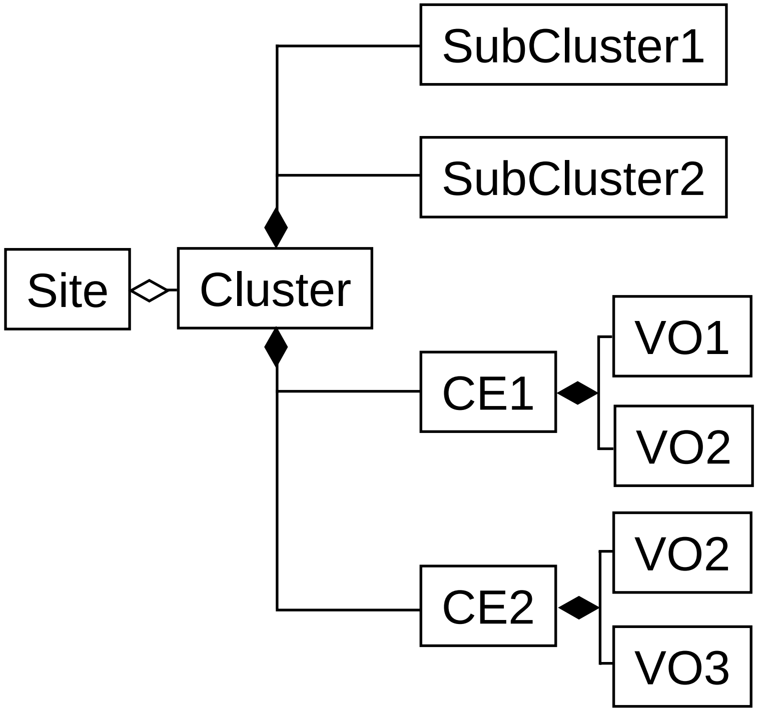
Glue Schema to old classad Mapping

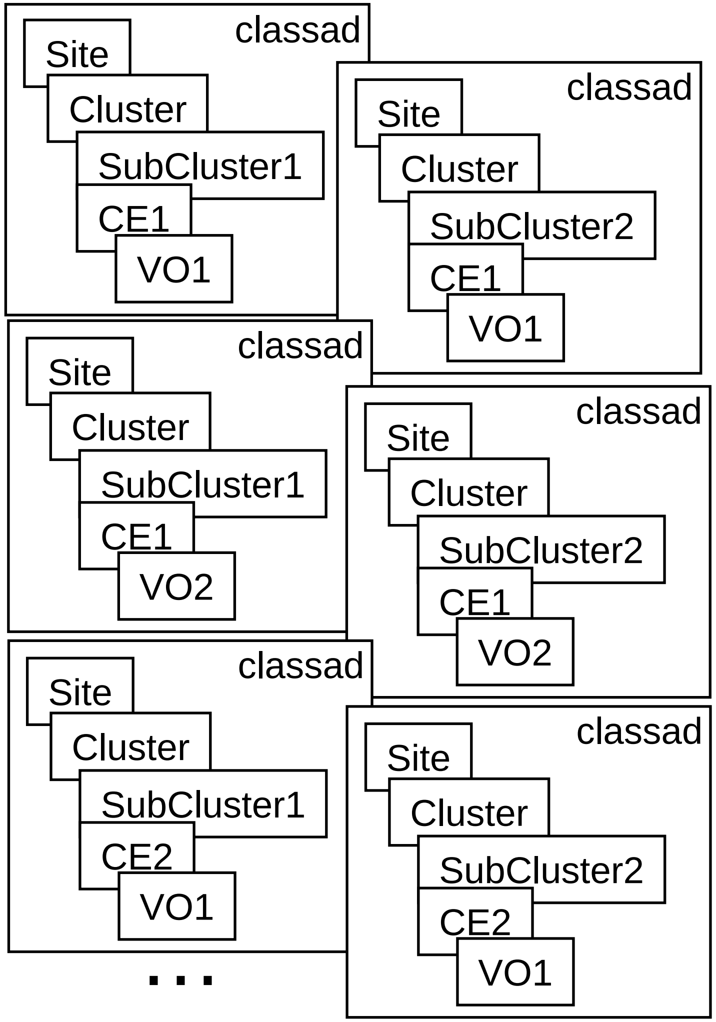


Mapping the Glue Schema “tree” into a set of “flat” classads:
All possible combination of (Cluster, Subcluster, CE, VO)

…

Glue Schema to old classad Mapping

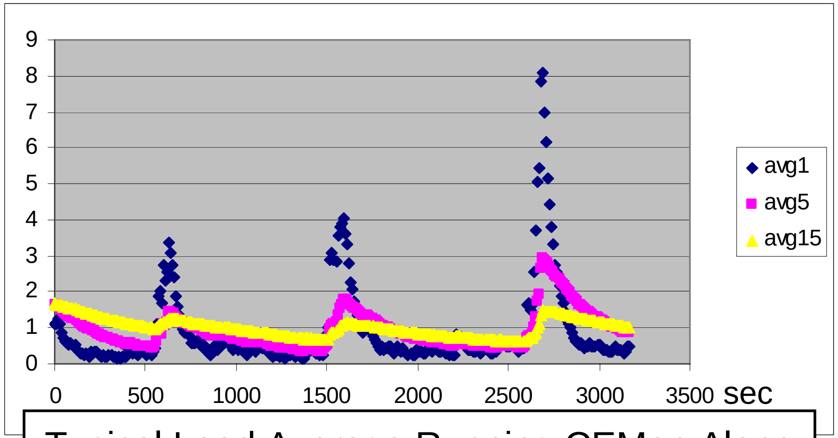




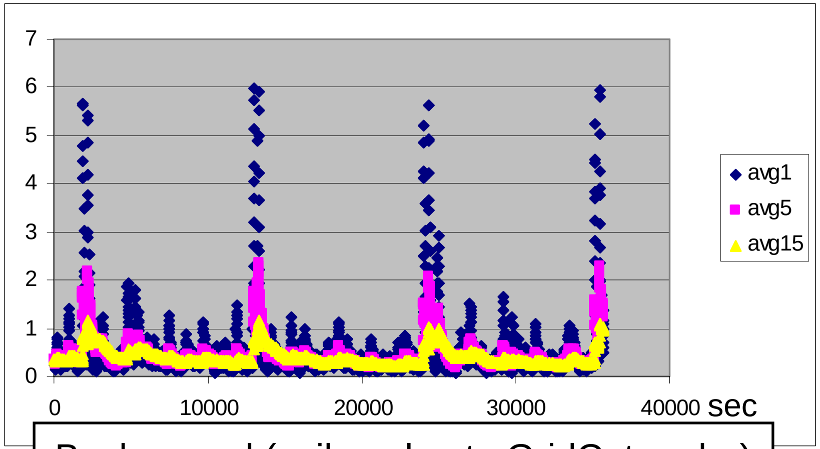
Mapping the Glue Schema “tree” into a set of “flat” classads:
All possible combination of (Cluster, Subcluster, CE, VO)

Impact of CEMon on the OSG CE

- **We studied CEMon resource requirements (load, mem, …) at a typical OSG CEs**
  - CEMon *pushes* information periodically
- **We compared CEMon resource requirements with MDS-2 by running**
  - CEMon alone (invokes GIP)
  - GRIS alone (Invokes GIP) queried at high-rate (many LCG Brokers scenario)
  - GIP manually
  - CEMon AND GRIS together
- **Conclusions**
  - **running CEMon alone does not generate more load than running GRIS alone or running CEMon and GRIS**
  - **CEMon uses less %CPU than a GRIS that is queried continuously (0.8% vs. 24%). On the other hand, CEMon uses more memory (%4.7 vs. %0.5).**
- More info at https://twiki.grid.iu.edu/twiki/bin/view/ResourceSelection/CEMonPerformanceEvaluation



Typical Load Average Running CEMon Alone

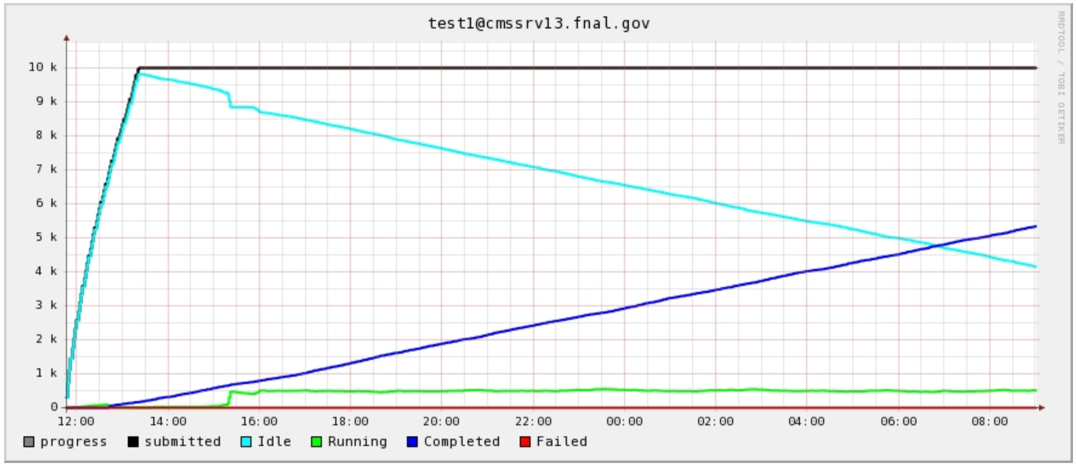


Background (spikes due to GridCat probe)

Open Science Grid

US CMS evaluates WMS’s

- Condor-G test with manual res. selection (NO ReSS)
  - Submit 10k sleep jobs to 4 schedulers
  - Jobs last 0.5 – 6 hours
  - Jobs can run at 4 Grid sites w/ ~2000 slots
- When Grid sites are stable, Condor-G is scalable and reliable

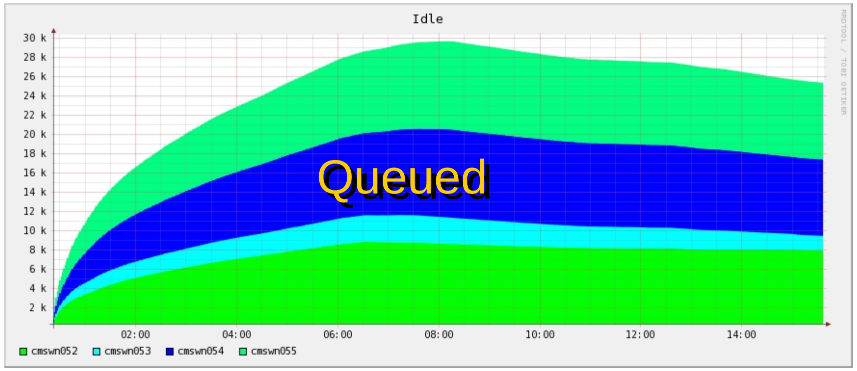


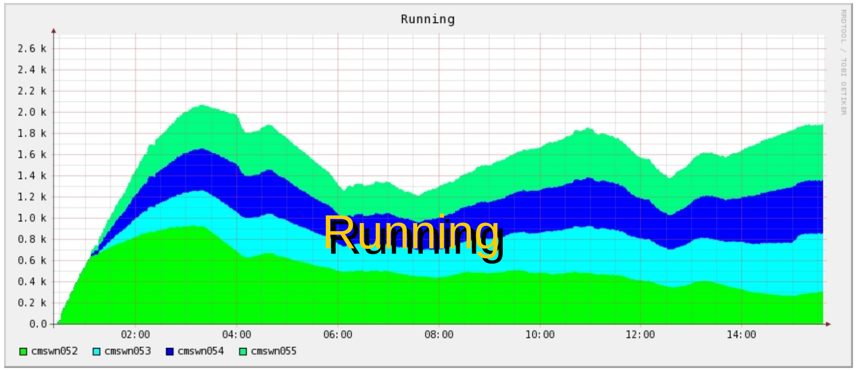
1 Scheduler view of Jobs Submitted, Idle, Running, Completed, Failed Vs. Time

Study by Igor Sfiligoi & Burt Holzman,
US CMS / FNAL, 03/07
https://twiki.grid.iu.edu/twiki/bin/view/ResourceSelection/ReSSEvaluationByUSCMS

ReSS Scalability

- Condor-G + ReSS Scalability Test
  - Submit 10k sleep jobs to 4 schedulers
  - 1 Grid site with ~2000 slots; multiple classad from VOs for the site
- Result: same scalability as Condor-G
  - Condor Match Maker scales up to 6k classads

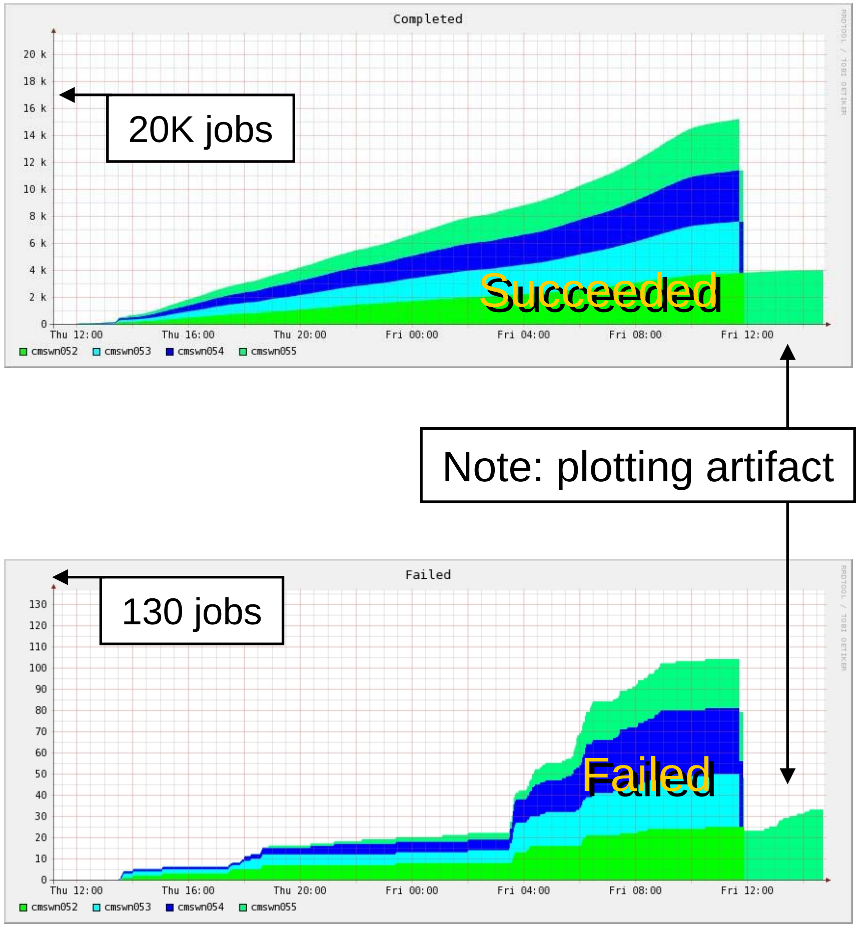




Open Science Grid

ReSS Reliability

- Same reliability as Condor-G, when grid sites are stable
- Failures mainly due to Condor-G / GRAM communication problems.
- Failures can be automatically resubmitted / re-matched (not tested here)



20K jobs

Note: plotting artifact

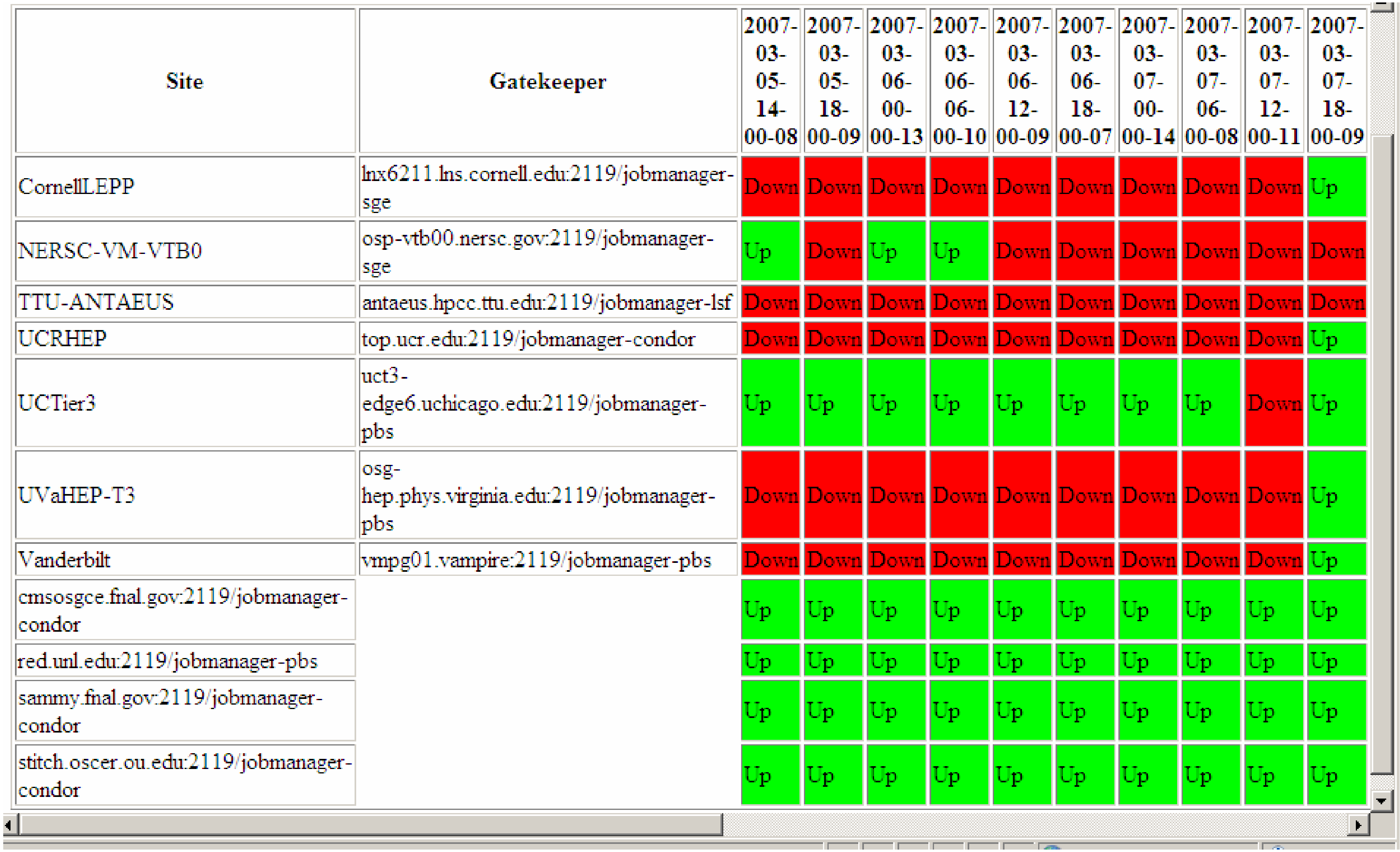
130 jobs

Open Science Grid

Project Status and Plans

- Development is mostly done
  - We may still add SE to the resource selection process
- ReSS is now the resource selector of Fermigrid
- Assisting Deployment of ReSS (CEMon) on Production OSG sites
- Using ReSS on SAM-Grid / OSG for DZero data reprocessing for the available sites
- Working with OSG VOs to facilitate ReSS usage
- Integrate ReSS with GlideIn Factory
- Move the project to maintenance

ReSS Deployment on OSG



Site	Gatekeeper	2007-03-05-14-00-08	2007-03-05-18-00-09	2007-03-06-00-00-13	2007-03-06-06-00-10	2007-03-06-12-00-09	2007-03-06-18-00-07	2007-03-07-00-00-14	2007-03-07-06-00-08	2007-03-07-12-00-11	2007-03-07-18-00-09
CornellLEPP	lnx6211.lns.cornell.edu:2119/jobmanager-sge	Down	Down	Down	Down	Down	Down	Down	Down	Down	Up
NERSC-VM-VTB0	osp-vtb00.nersc.gov:2119/jobmanager-sge	Up	Down	Up	Up	Down	Down	Down	Down	Down	Down
TTU-ANTAEUS	antaeus.hpcc.ttu.edu:2119/jobmanager-lsf	Down	Down	Down	Down	Down	Down	Down	Down	Down	Down
UCRHEP	top.ucr.edu:2119/jobmanager-condor	Down	Down	Down	Down	Down	Down	Down	Down	Down	Up
UCTier3	uct3-edge6.uchicago.edu:2119/jobmanager-pbs	Up	Up	Up	Up	Up	Up	Up	Up	Down	Up
UVaHEP-T3	osg-hep.phys.virginia.edu:2119/jobmanager-pbs	Down	Down	Down	Down	Down	Down	Down	Down	Down	Up
Vanderbilt	vmpg01.vampire:2119/jobmanager-pbs	Down	Down	Down	Down	Down	Down	Down	Down	Down	Up
cmsosgce.fnal.gov:2119/jobmanager-condor		Up	Up	Up	Up	Up	Up	Up	Up	Up	Up
red.unl.edu:2119/jobmanager-pbs		Up	Up	Up	Up	Up	Up	Up	Up	Up	Up
sammy.fnal.gov:2119/jobmanager-condor		Up	Up	Up	Up	Up	Up	Up	Up	Up	Up
stitch.oscer.ou.edu:2119/jobmanager-condor		Up	Up	Up	Up	Up	Up	Up	Up	Up	Up

Click here for live URL

Open Science Grid

Conclusions

- ReSS is a lightweight Resource Selection Service for push-based job handling systems
- ReSS is deployed on OSG 0.6.0 and used by FermiGrid
- More info at http://osg.ivdgl.org/twiki/bin/view/ResourceSelection/