Where do I send my jobs?

Grid information systems in the OSG

Grid Information Systems

- What is the grid? A physicist once told me this:
  - "My original idea of a grid was the power grid; just like you plug a lamp into the wall without worrying what powerplant the electricity came from, you should be able to send your jobs to the grid without worrying where the CPU is coming from."
  - Not quite….

Grids, In General

- It turns out that, just looking at computation, there are many differences.
- We originally hoped all plugs look like this:

The Grid

- It turns out, there are many important differences!

Information Systems

- Grid Information Systems allow you to describe what kind of grid jobs you accept.
  - Then, a user can describe what kind of resources their job works with.
  - And then a matchmaker will try to match a job to a computer.

Grid Information Systems

- You say you have this:

- I say I have this:

- And then hopefully this happens:

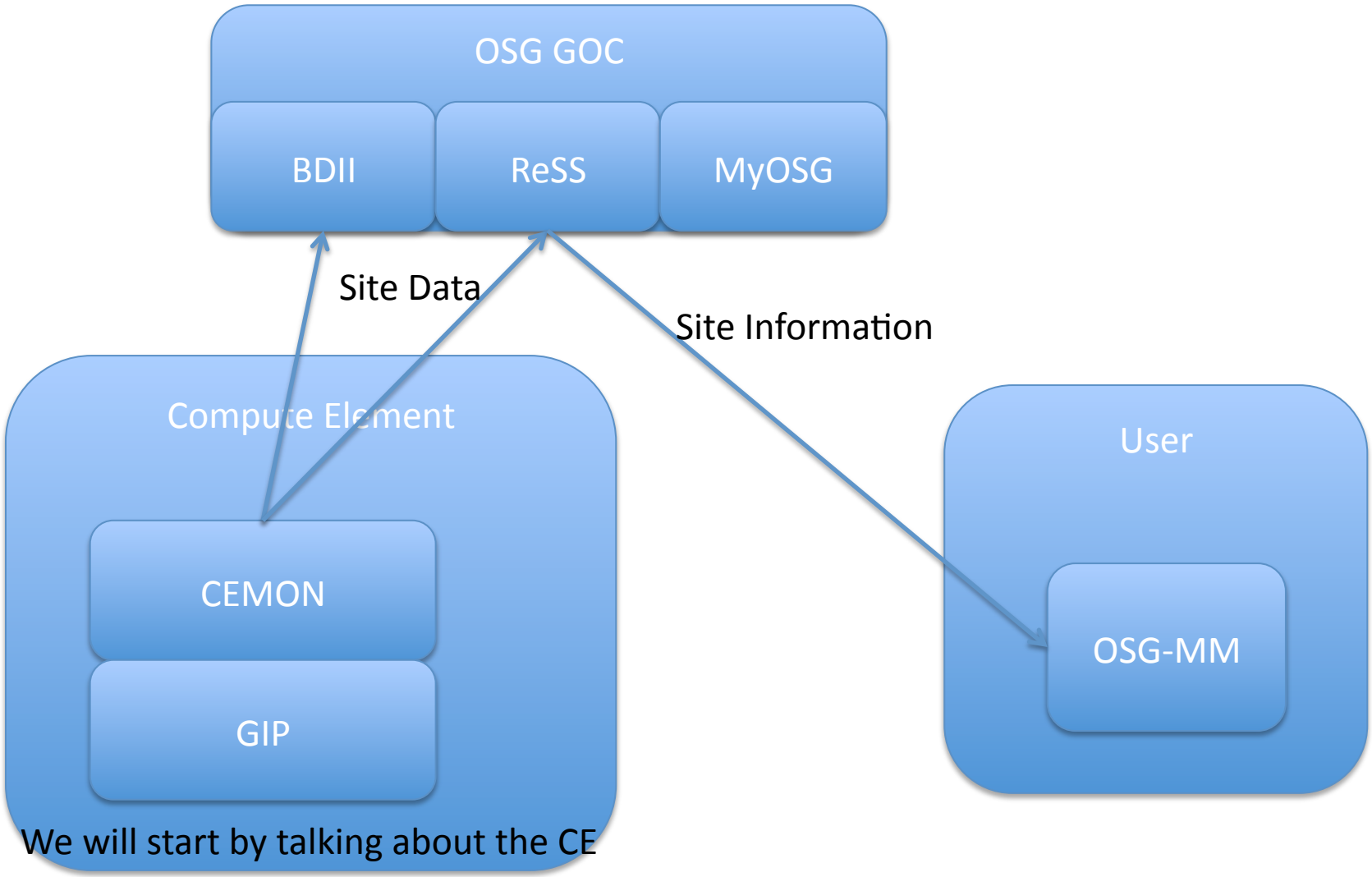
Grid Information Systems

- Of course, sometimes this happens

In the OSG

- In the OSG, this is what we call our various pieces:
  - On the compute element: GIP, CEMon
  - Centrally at the GOC or FNAL: IG/ReSS, BDII
  - User tools: ldapsearch, OSG-MM, Pegasus, storage discovery tools
- In this presentation, we'll talk about the GIP, CEMon, ReSS, and BDII. I believe the client-side tools will be covered by others.

The “Big Picture”



Info Services on the CE

- The OSG CE runs two pieces of software for information services:
  - GIP (Generic Information Provider)
  - CEMon
- The GIP queries various components of the CE (and maybe the associated SE) and comes up with a description of the CE.
  - The description is in a schema called “GLUE” and written as LDIF.

GLUE Schema

- The GLUE schema is the heart of the OSG Information Services.
  - It is a schema that defines the way to describe your cluster.
  - This “description” can be written in several ways – XML and LDAP are most popular.
  - A schema is important; goal is to unambiguously describe a grid site independent of what technology the site is using.

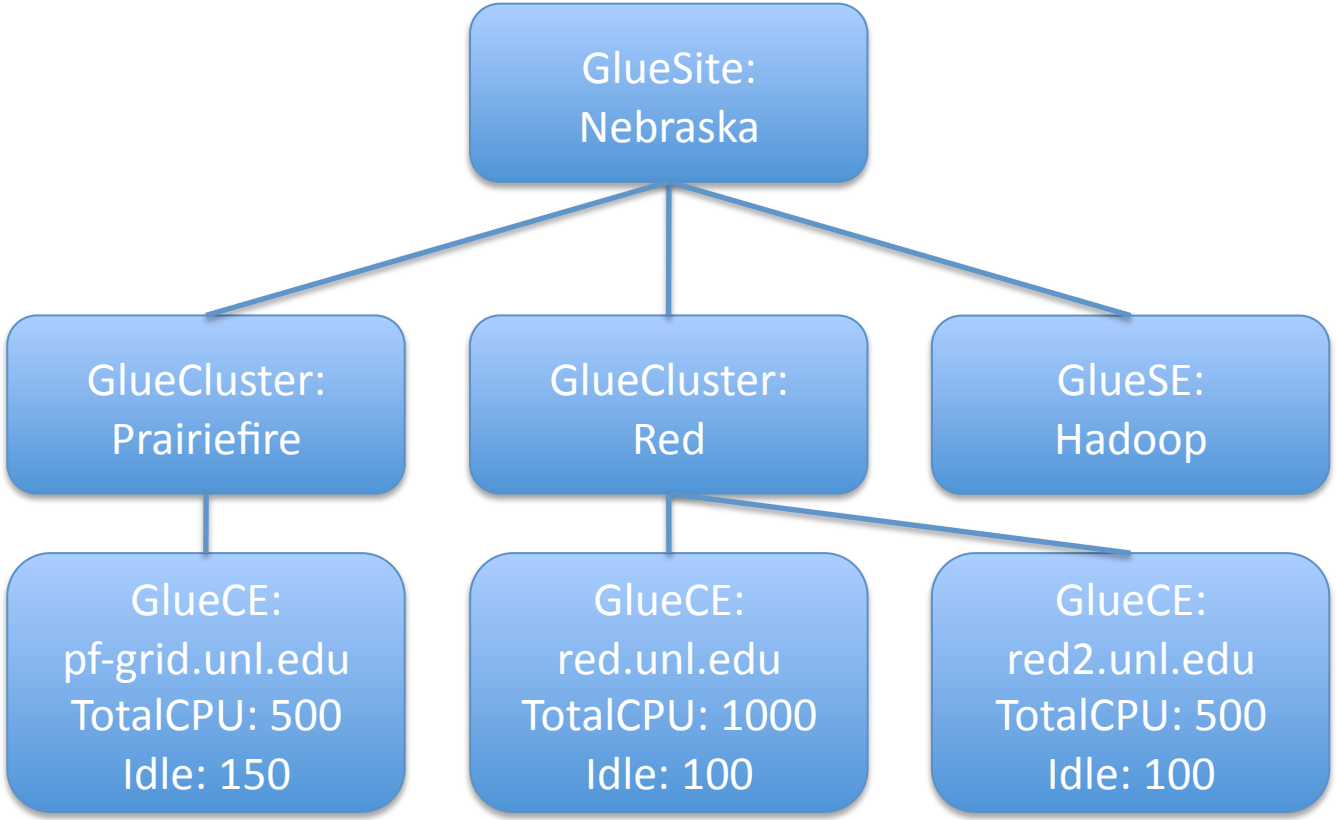
GLUE Schema

- Because it aims to be able to describe any grid site in the world, GLUE is *quite complex*.
  - I hope an end user *never has to read GLUE directly.*
  - A user *should* understand the data model though.
- These are the primary concepts:
  - Site. A collection of clusters and storage.
  - Cluster/Subcluster. A cluster is a collection of computers under a batch scheduler; a subcluster is a collection of computers running the same hardware.
  - Compute Element (CE). A grid gateway into the cluster; represents a queue/gatekeeper combo.
  - VOView. Information about a single VO's activities on a CE.
  - Storage Element (SE). A system that stores data.
  - Storage Area. A logical area in the SE.

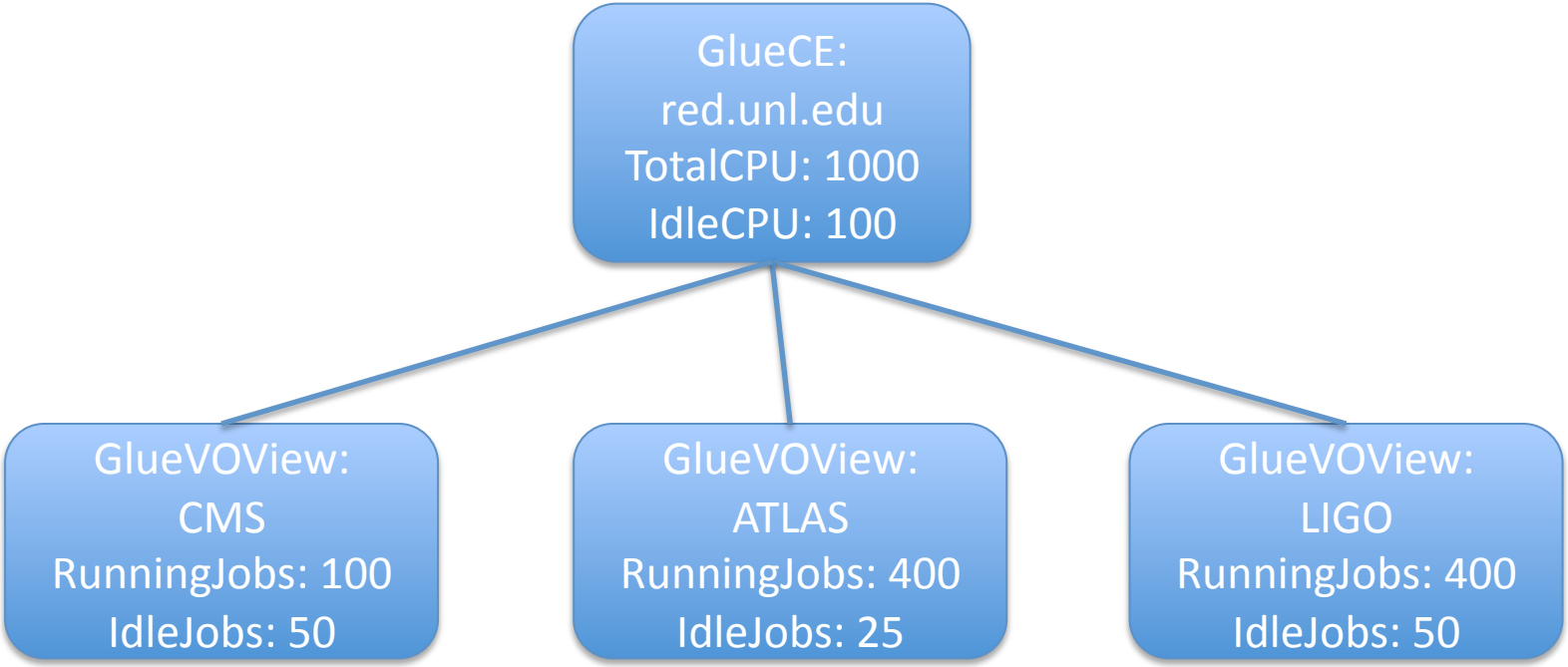
Example

- Let's say you have a site named "Nebraska" (that's my site!)
- And two clusters named "Prairiefire" and "Red".
- Red has a storage element called "Hadoop"
- Red and Prairiefire both run Condor.
  - Red has 2 OSG CEs: red.unl.edu and red2.unl.edu
  - Prairiefire has 1 OSG CE: pf-grid.unl.edu
- Red has 1000 CPUs, 100 idle. Prairiefire has 500 CPUs, 150 idle.
- How do you describe this site?

GLUE Example



GLUE Example



Real-Life GLUE

dn: GlueCEUniqueID=cit-gatekeeper.ultralight.org:2119/jobmanager-condor-cms_production,mds-vo-name=CIT_CMS_T2,mds-vo-name=local,o=grid
objectClass: GlueCE
objectClass: GlueCEAccessControlBase
objectClass: GlueCEInfo
objectClass: GlueCEPolicy
objectClass: GlueCEState
objectClass: GlueCETop
objectClass: GlueInformationService
objectClass: GlueKey
objectClass: GlueSchemaVersion
GlueCEInfoDataDir: /raid2/osg-data
GlueCEPolicyMaxObtainableCPUTime: 1440
GlueCEStateRunningJobs: 140
GlueSchemaVersionMajor: 1
GlueCEInfoTotalCPUs: 350
GlueCEStateFreeJobSlots: 2
GlueCEPolicyMaxWaitingJobs: 99999
GlueCEStateWorstResponseTime: 261651
GlueCEPolicyMaxTotalJobs: 99999
GlueCEPolicyMaxObtainableWallClockTime: 1440
GlueCEStateTotalJobs: 567
GlueCEStateStatus: Production
GlueForeignKey: GlueClusterUniqueID=caltech-cms-t2
GlueCECapability: CPUScalingReferenceSI00=2000
GlueCEAccessControlBaseRule: VO:cms
GlueCEInfoLRMSType: condor
GlueCEPolicyMaxRunningJobs: 2000
GlueCEPolicyAssignedJobSlots: 350
GlueCEInfoApplicationDir: /raid1/osg-app
GlueCEPolicyPreemption: 0
GlueCEStateFreeCPUs: 2
GlueCEInfoGRAMVersion: 2.0
GlueCEImplementationName: Globus
GlueSchemaVersionMinor: 3
GlueCEStateEstimatedResponseTime: 40992
GlueCEHostingCluster: cit-gatekeeper.ultralight.org
GlueCEInfoHostName: cit-gatekeeper.ultralight.org
GlueCEInfoDefaultSE: cit-se.ultralight.org
GlueCEImplementationVersion: 4.0.6
GlueCEInfoLRMSVersion: 7.2.0 Dec 19 2008 BuildID: 121001 $

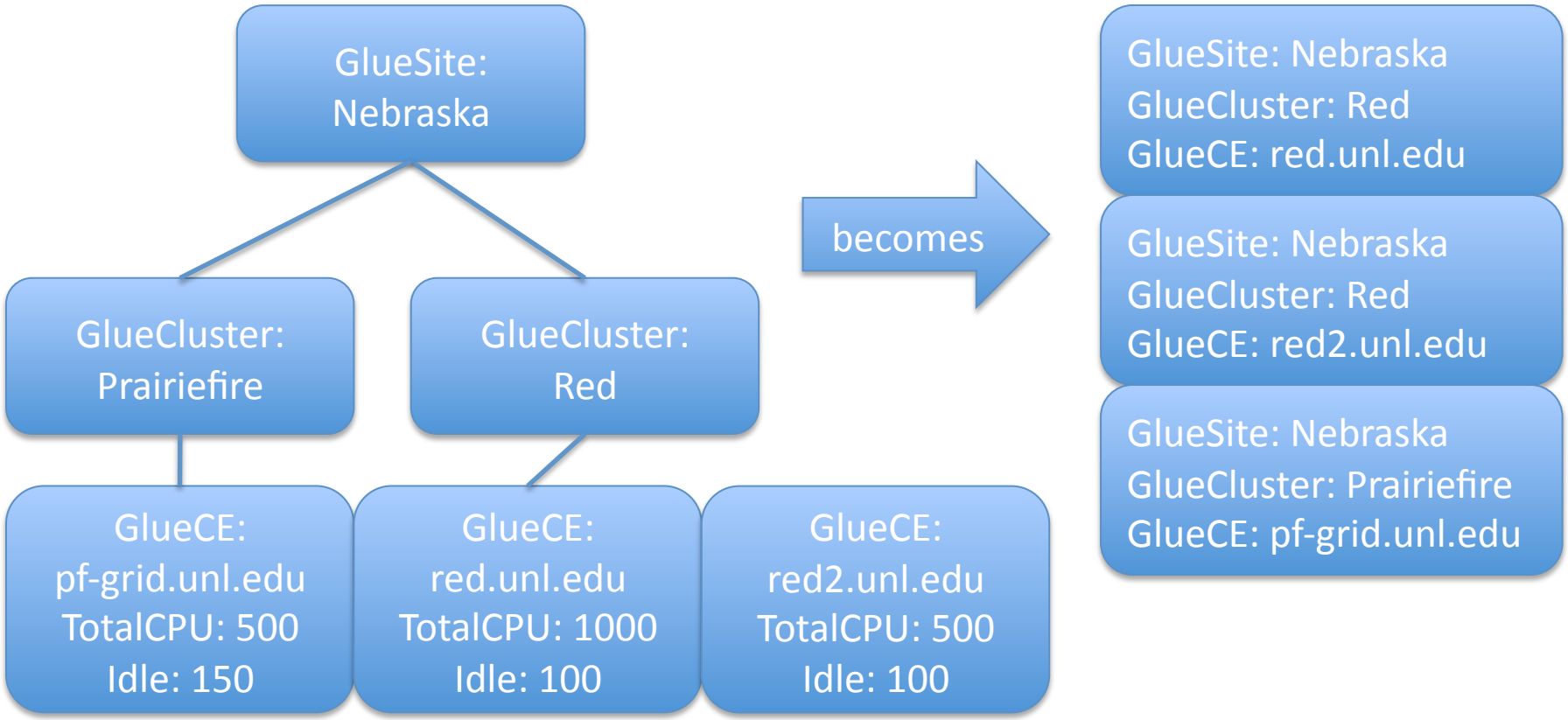
- These abstract ideas are often represented using LDAP. Here's part of an example GlueCE; you'll see why it's not user-friendly.

That's nice, now what?

- Again, the job of the GIP is to automatically create the GLUE description of the site using the information found in the OSG CE.
- CEMon is a web application that runs the GIP every 5 minutes (like cron).
- CEMon takes the LDIF output of the GIP and sends it to the central OSG servers.
  - One server gets the raw LDIF
  - Another gets the output transformed into Condor ClassAds

Condor ClassAds

- Instead of having a hierarchy of LDIF stanzas to analyze, CEMon creates one stanza per combo



CEMon

- CEMon sends the raw BDII to the GOC.
  - The destination CEMon at GOC feeds this into LDAP servers.
- CEMon sends the Condor ClassAds to FNAL.
  - This destination CEMon places them into a central Condor scheduler.

From OSG to you

- The largest user of the raw LDIF is CMS as this is used to power the gLite WMS. The GIP/CEMon infrastructure allows CMS sites in the US to interoperate with CMS sites in Europe.
- Because Condor-G is the most popular tool (maybe de-facto) to submit jobs on the grid, the Condor ClassAds are easiest to use.
  - OSG-MM, for example, imports all ClassAds that advertise support for your VO into a local Condor instance.
  - These ClassAds are then ranked using test jobs.
  - And you are left with a list of usable, well-behaved sites you can run against.

OSG-MM

- My site is a happy user of OSG-MM. Our grad student has used it to run almost 1M hours of jobs in the last month.
- Can't emphasize strongly enough the *ease of use*. Here's a sample submit file:

```
transfer_input_files = srm://srm.unl.edu:8446/srm/v2/server?SFN=/mnt/hadoop/user/dweitzel/gpn/
USCensus1990.data.txt.no1line
executable = sampleCreator.py
training = 100000
testing = 30
Args = USCensus1990.data.txt.no1line $(training) $(testing)

Rank = (TARGET.Rank) + TARGET.GlueCEStateFreeCPUs
queue 20
```

Other Related Tools

- I generally think there are three broad categories of information:
  - Site topology (OSG CE on host X is related to OSG SE on host Y).
  - State of site's batch system (site is 95% occupied)
  - Monitoring status of site (Globus appears to be working).
- The GIP is generally good for the first two, but because it is located on the CE itself, it is not useful for monitoring the CE health.

MyOSG

- MyOSG is an easy-to-use web interface to a lot of the OSG information.
  - It is the official record of OSG topology
  - It reports the results of RSV, the OSG's testing tool. This provides health information for sites.
- MyOSG *does* provide machine-readable information about CE/SE status.
- It *does not* provide machine-readable to submit to Globus or use a SE.
  - Yet, it is one of my favorite tools because it allows you to explore the various services easily.

About MyOSG – MyOSG

iu.edu https://myosg.grid.iu.edu/about

Most Visited Getting Started Latest Headlines Apple Amazon eBay Yahoo! News

Grid Columbia Workshop (20–3... Results of Query: Cumulative H... About MyOSG – MyOSG

MyOSG
by Grid Operations Center

Home | Resource Group | Support Center | Virtual Organization | Status Map | Miscellaneous

About MyOSG

MyOSG is designed with the primary goal of providing users, administrators, VO managers, and everyone else, a one-stop location for various pieces of OSG related information.

MyOSG allows you to quickly select and filter information you are looking for. Most pages also allow you to export the selected/filtered data in their preferred format HTML, Netvibes, Google, etc. and XML for programmatic interfaces.

Quick Links

Resource Summary
Current status summary from all production OSG resources (*ITB Resource Summary*).

RSV Status History
Last 7 days worth of RSV-based status history for all production resources. You can click on any point (in time) on the graph to display the metric details at that time instant (*ITB RSV Status History*).

Current RSV Status
Current RSV-based status for all production resources (*Current RSV Status for ITB resources*).

GIP Validation Results
Current GIP validation status for all production resources that provide the CE service (*GIP validation results for ITB resources*).

Gratia Accounting
Gratia Accounting based CPU usage information grouped by username for the last 30 days for all production resources (*Gratia Accounting graph for ITB resources*).

Status Map
"Status Map" shows the current overall *site* status on a Google Maps based world map (*Status Map for ITB sites*).

Externa

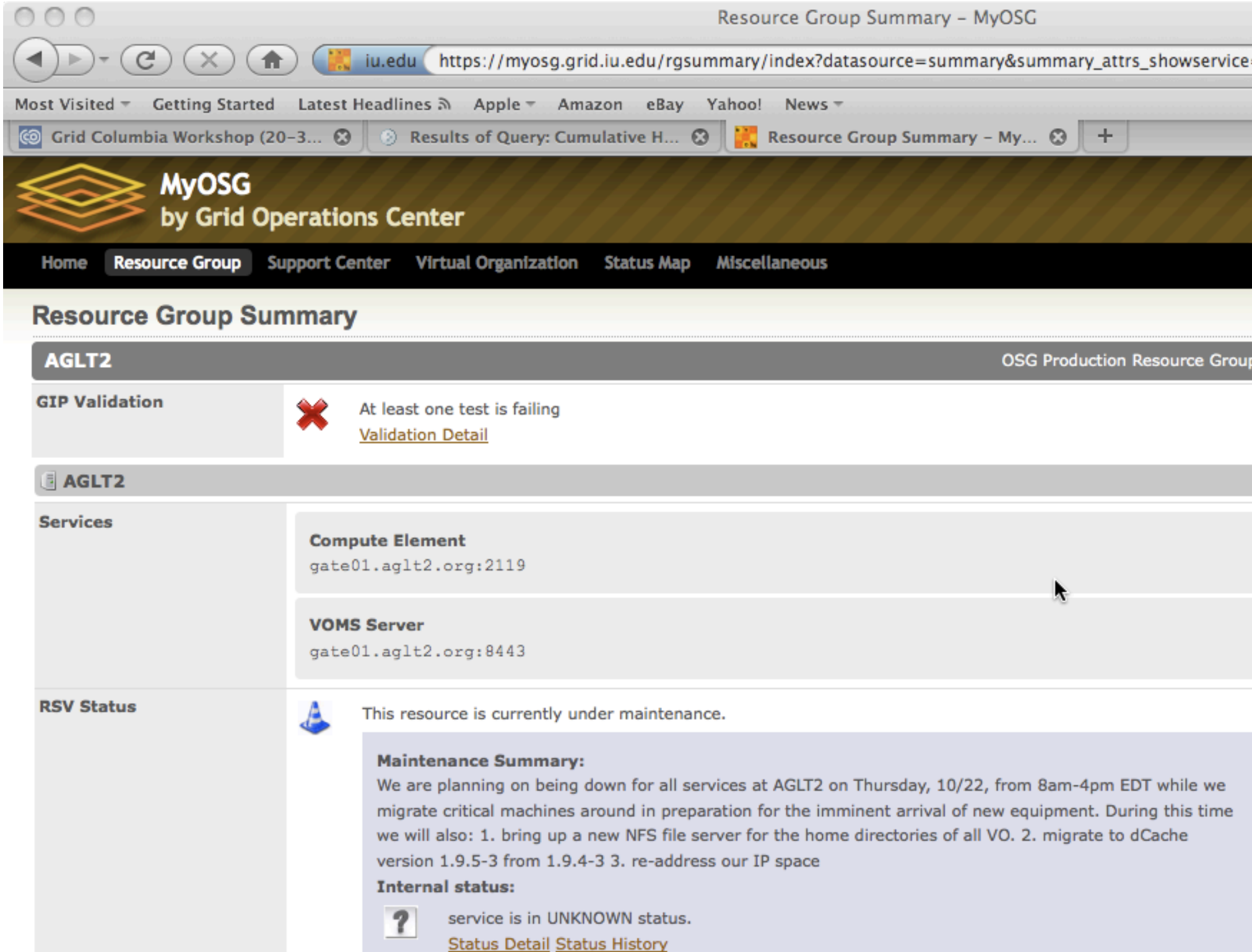
OSG I

virtual or

GOC C

OSG T

and maili
are open
area com

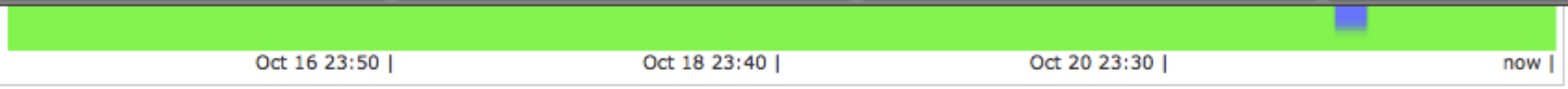

Resource Group Summary – MyOSG
iu.edu
https://myosg.grid.iu.edu/rgsummary/index?datasource=summary&summary_attrs_showservice
Most Visited
Getting Started
Latest Headlines
Apple
Amazon
eBay
Yahoo!
News
Grid Columbia Workshop (20–3...
Results of Query: Cumulative H...
Resource Group Summary – My...
MyOSG
by Grid Operations Center
Home
Resource Group
Support Center
Virtual Organization
Status Map
Miscellaneous
Resource Group Summary
AGLT2
OSG Production Resource Group
GIP Validation
At least one test is failing
Validation Detail
AGLT2
Services
Compute Element
gate01.aglt2.org:2119
VOMS Server
gate01.aglt2.org:8443
RSV Status
This resource is currently under maintenance.
Maintenance Summary:
We are planning on being down for all services at AGLT2 on Thursday, 10/22, from 8am-4pm EDT while we migrate critical machines around in preparation for the imminent arrival of new equipment. During this time we will also: 1. bring up a new NFS file server for the home directories of all VO. 2. migrate to dCache version 1.9.5-3 from 1.9.4-3 3. re-address our IP space
Internal status:
service is in UNKNOWN status.
Status Detail Status History

RSV Status History – MyOSG

iu.edu https://myosg.grid.iu.edu/rgstatushistory/index?datasource=statushistory&account_type=cumulative_hou

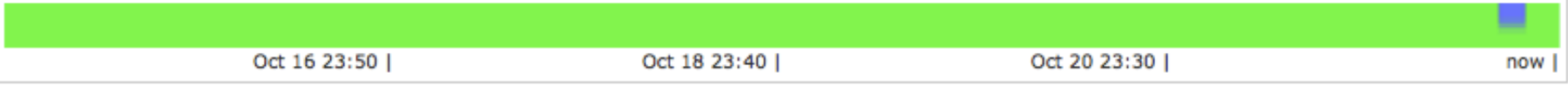
Most Visited Getting Started Latest Headlines Apple Amazon eBay Yahoo! News

Grid Columbia Workshop (20–3... Results of Query: Cumulative H... Gratia Accounting – Daily Hour... RSV Status History – MyO



CIT_CMS_T2B

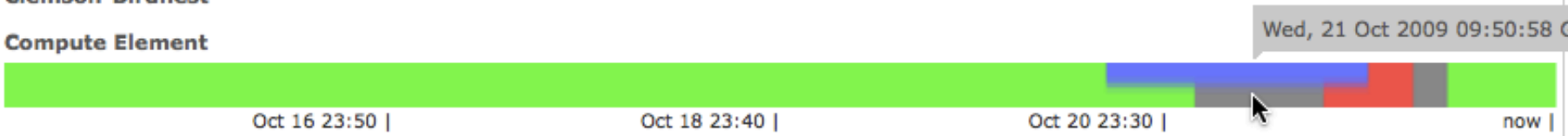
Compute Element



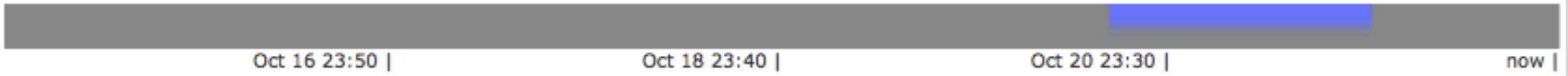
Clemson-Birdnest — OSG Production Resource Group

Clemson-Birdnest

Compute Element



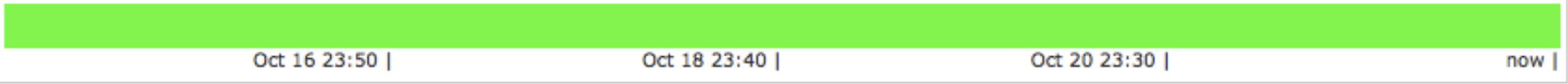
SRM V2 Storage Element



Clemson-ciTeam — OSG Production Resource Group

Clemson-ciTeam

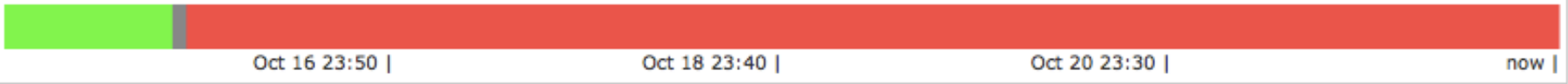
Compute Element



CLEMSON-Palmetto — OSG Production Resource Group

CLEMSON-Palmetto

Compute Element



CNIC-CIGI-OSG — OSG Production Resource Group

Wrap-up

- Now that you know how to submit jobs to a site, you need to know *what site* to submit to.
  - Such information can be the difference between a pleasant and unpleasant experience.
- It is the job of the Grid Information Services to provide you with the data you need to make this decision.
- The “language” Grid Information Services uses is called GLUE; it is a data model to describe one’s site, but only experts want to use it directly.

Wrap-up

- This information starts out at the site in many components.
  - The GIP gathers it and converts it to GLUE.
  - CEMon is responsible for sending and transforming the GLUE to central services.
- Central services provide a view of the entire grid.
- Various client tools (such as OSG-MM) select the sites you care about and filter the results into something more usable.
  - The result is you can use the same tools (Condor-G) to submit your jobs and *mostly* forget about the hardworking grid information layers underneath.

Links

- OSG-MM: http://osgmm.sourceforge.net/
- MyOSG: https://myosg.grid.iu.edu/
- GLUE Working Group: http://forge.gridforum.org/sf/projects/glue-wg (all the technical details about GLUE)
- Questions about grid services? You can find us at osg-gip at opensciencegrid.org