Condor Week 2011

Why SaaS can be good

The tale of the OSG glidein factory

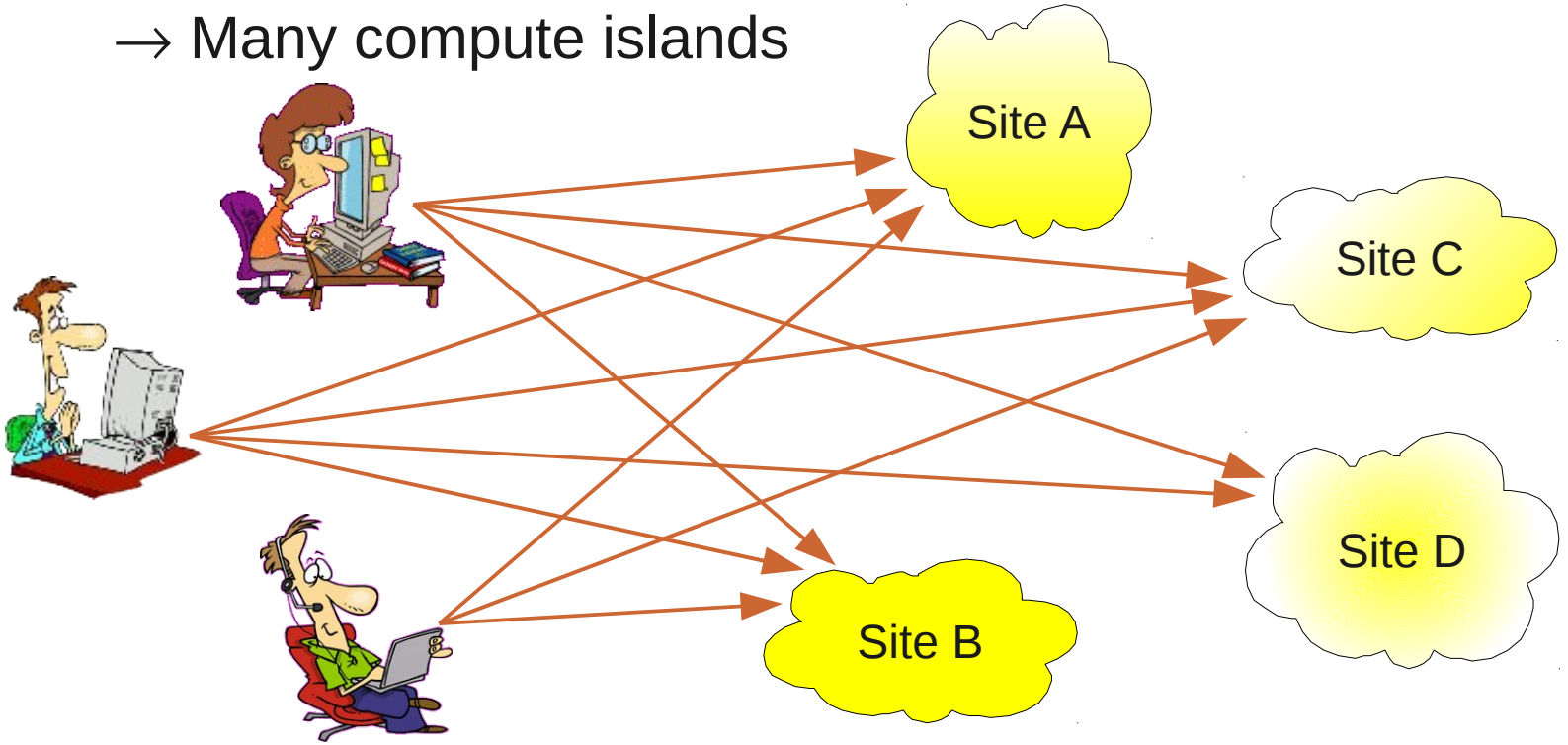
by Igor Sfiligoi and Jeff Dost
University of California San Diego

What are we talking about?

- We will try to convince you that the Grid experience can be much more pleasant if you use the path traced by the glideinWMS
  - We would love if you used glideinWMS itself, but that's not the main point
  - We are promoting the underlying principle
- And we will do it by providing some real life examples, too!

Some background -The grid

- Based on the principle of administrative autonomy
  - → Many compute islands

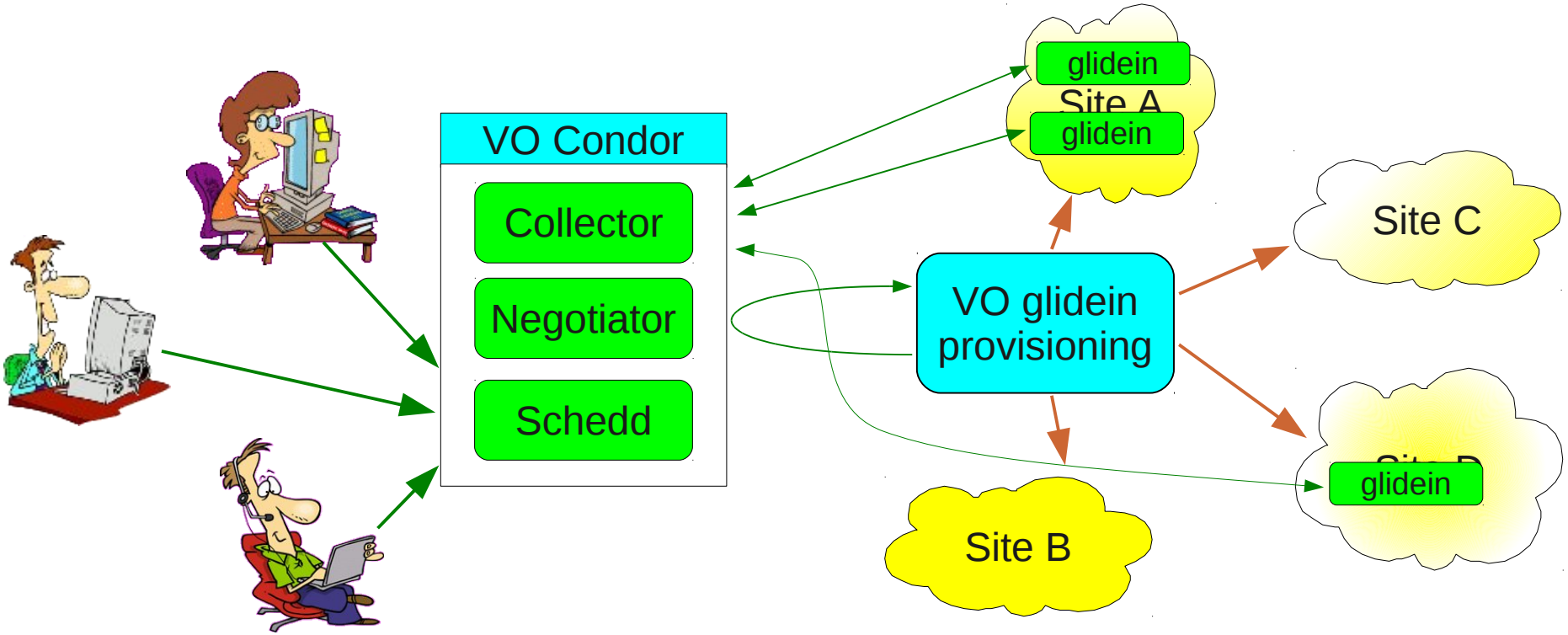


- Users have to handle errors from O(N) sources

Glideins make things better

for the users

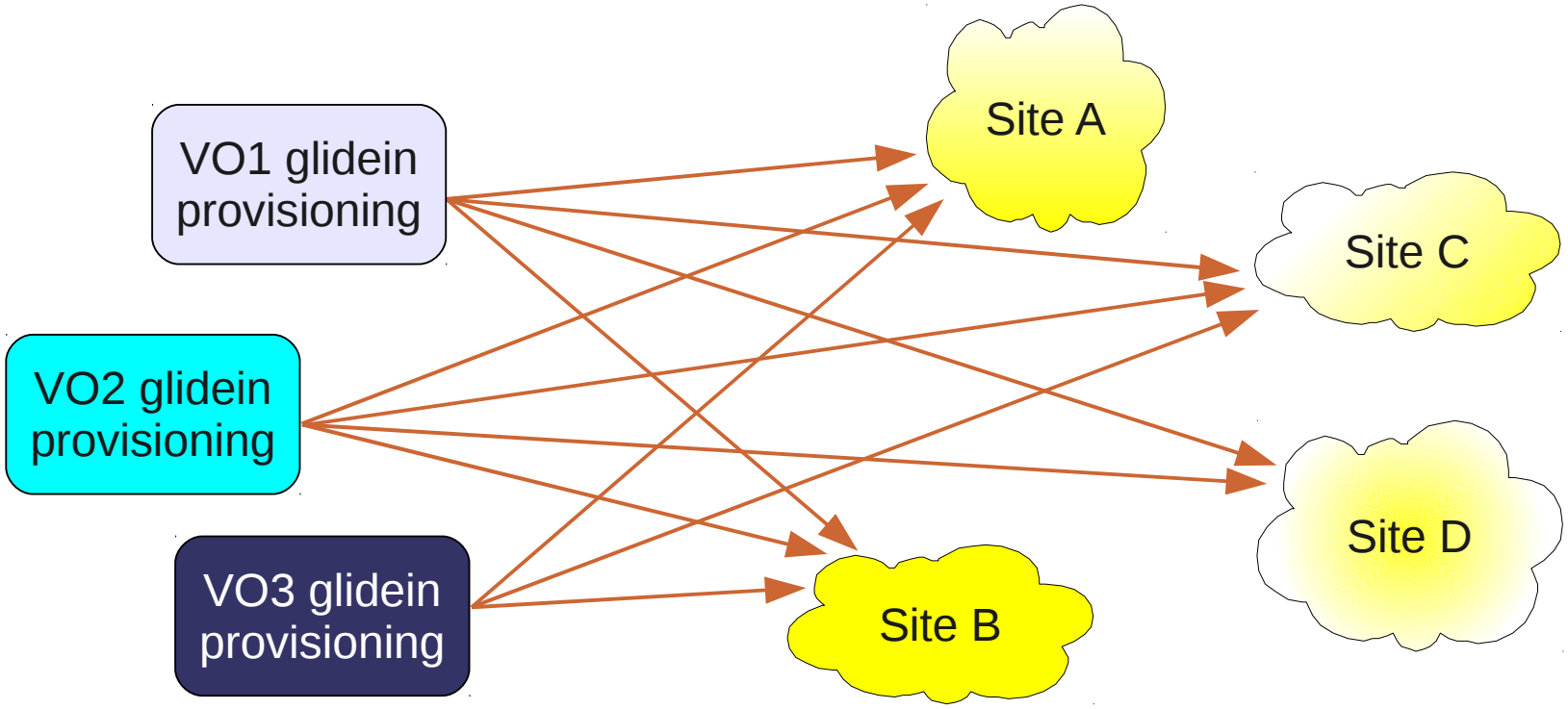
- Looks like a single Condor pool to users



- But more work for the VO admins
  VO = Virtual Organization (e.g. group)

Still N-to-M

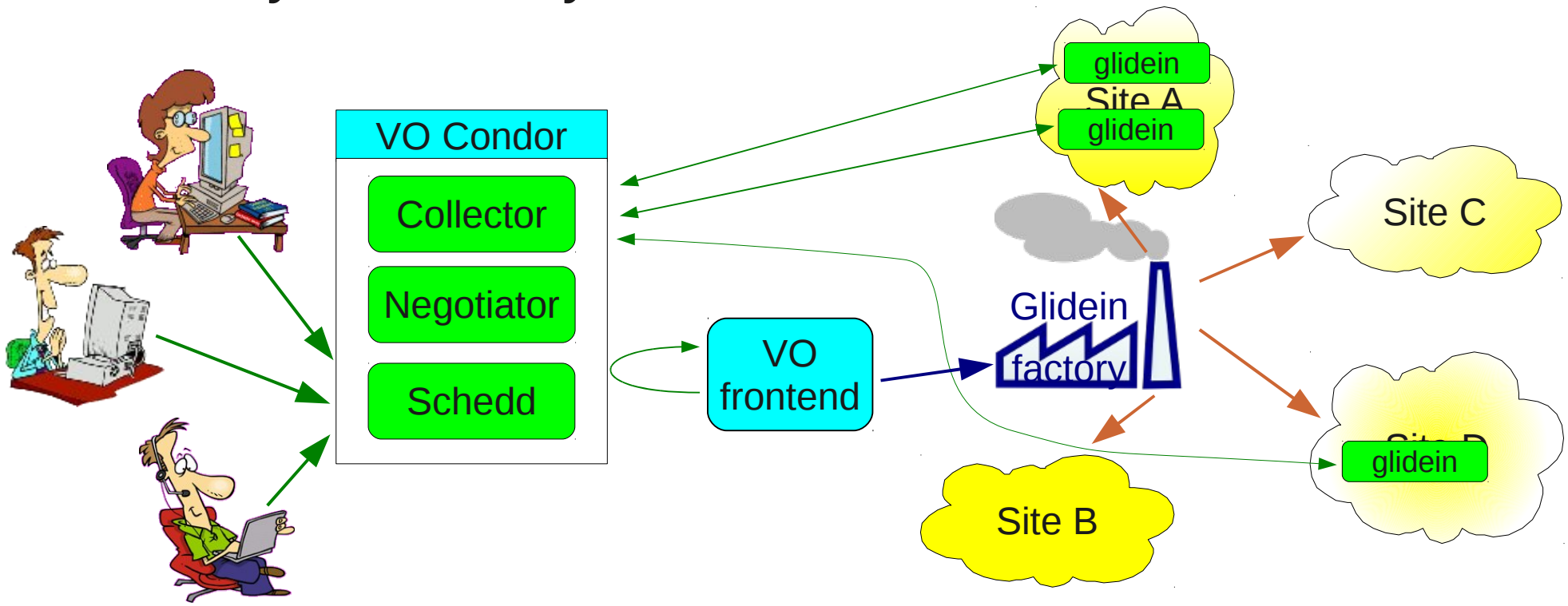
- Looks remarkably similar to initial problem



- Of course a few orders of magnitudes less entities

glideinWMS gets a step further

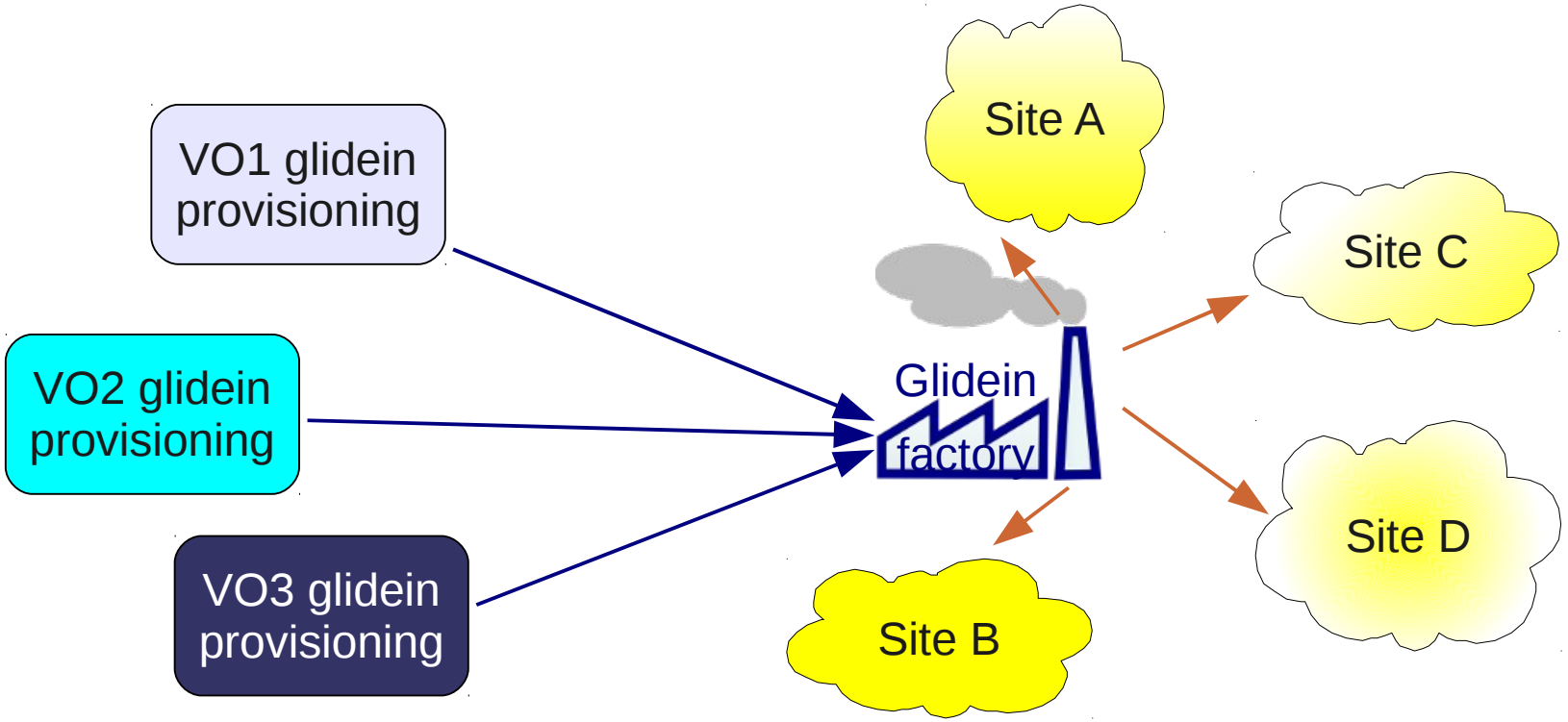
- Separates glidein submission logic from actual Grid submission of glideins
  - Only the factory sees the Grid



In glideinWMS factory can be shared

Although it does not need to be

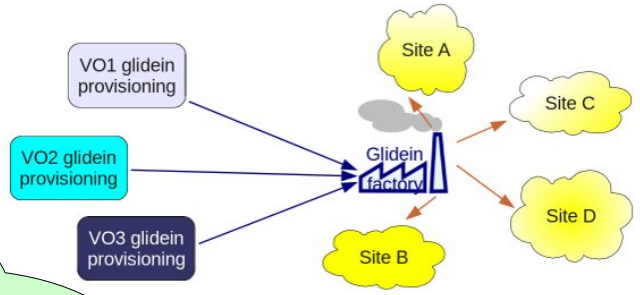
- Making life easy (also) for VO admins



- This (of course) means more work for the factory admins
  - But the promise is to lower the global cost

Enter SaaS

- Not a technology problem
  - Cannot be solved by software alone
- Someone needs to operate the service
  - The Global Factory

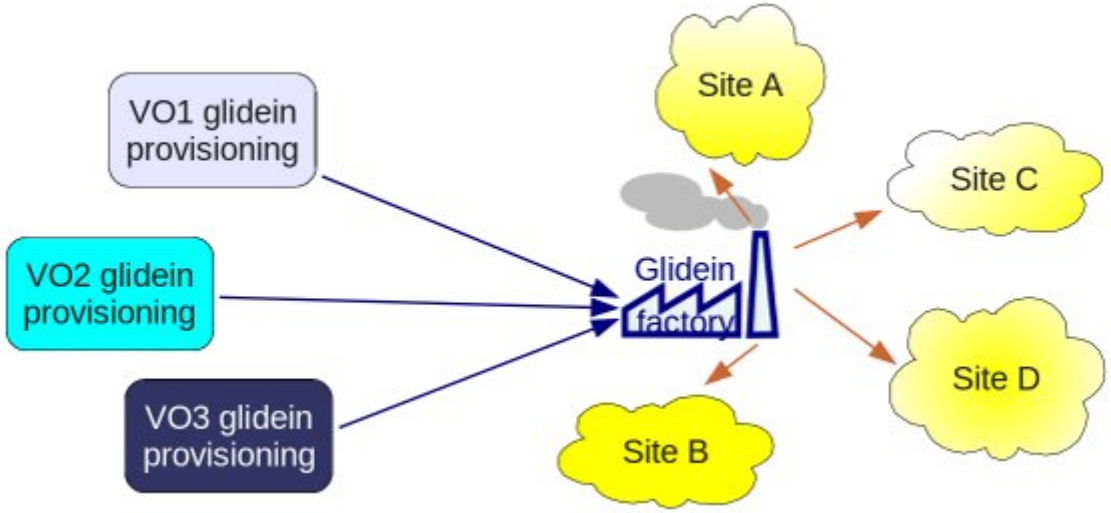


Usually referred to as

Software as a Service

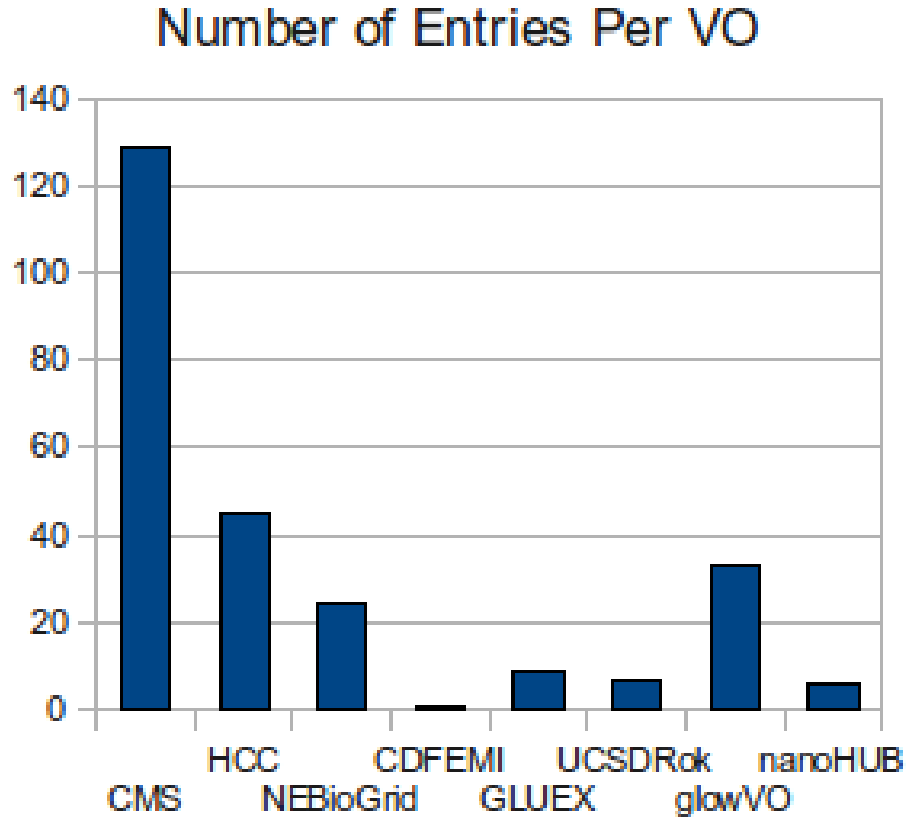
The OSG glidein factory

- Open Science Grid is a US Grid organization
  - Co-founded by NSF and DOE
- OSG is funding a glidein factory at UCSD
  - Open to all OSG VOs using glideinWMS frontends
  - Submitting glideins to both OSG and overseas sites



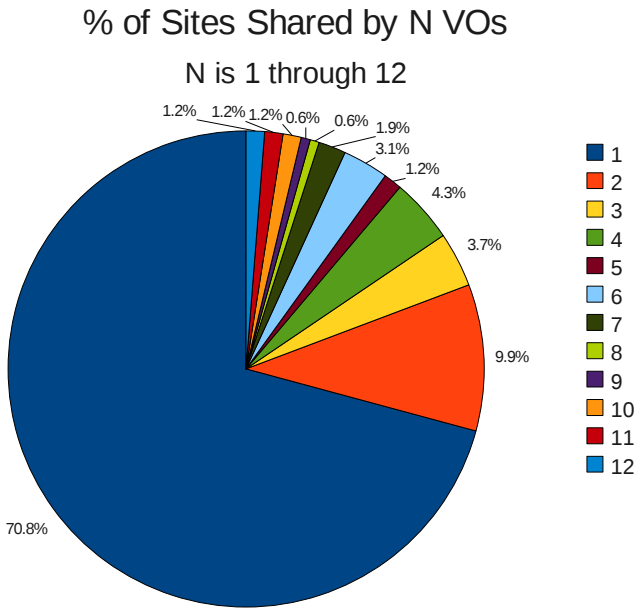
UCSD Factory Statistics

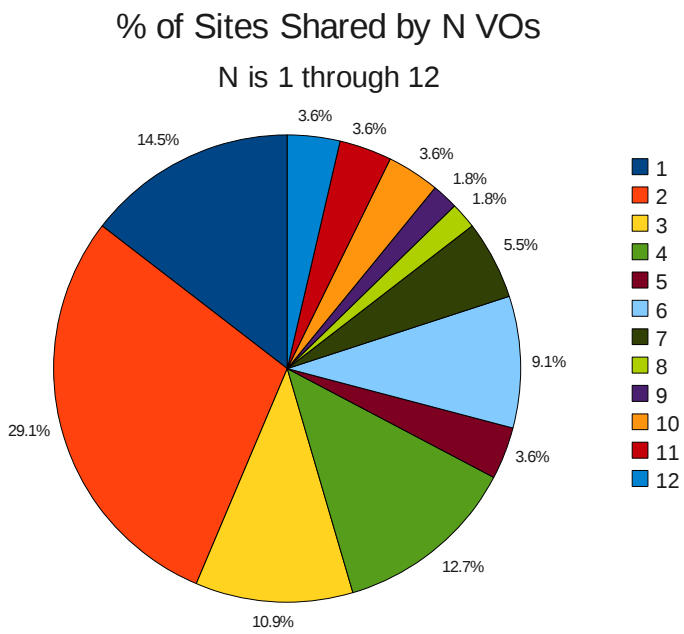
- ~10 active VOs served
- 160 entries total
- Many entries shared between VOs
- Biggest share
  - 132 CMS sites
- Not just OSG sites
  - 94 European CMS sites



UCSD Factory Statistics

- VO frontend site sharing





- Including CMS-only sites
- Excluding CMS-only sites

Configuration

- Create site entries for frontends to submit to
  - Add the entry to the factory configuration if it doesn't already exist
  - Add frontend VO names to entry whitelist to enable them to request glideins
- Monitor existing sites
  - Make factory config changes if site configs change
  - Temporarily stop submission if site is down for maintenance
- Site management from the factory keeps configurations centralized and out of the VO's hands

Factory Config File

```
<entry name="CMS_T2_US_UCSD_gw2" enabled="True" gatekeeper="osg-gw-2.t2.ucsd.edu:2119/jobmanager-
condor" gridtype="gt2" rsl="(queue=cms)(jobtype=single)" schedd_name="schedd_glideins20@glidein-
1.t2.ucsd.edu" verbosity="std" work_dir="Condor">
   <config>
      <max_jobs held="100" idle="400" running="10000"/>
      <release max_per_cycle="20" sleep="0.2"/>
      <remove max_per_cycle="5" sleep="0.2"/>
      <submit cluster_size="10" max_per_cycle="100" sleep="0.2"/>
   </config>
   <downtimes/>
   <allow_frontends>
   </allow_frontends>
   <attrs>
      <attr name="CONDOR_OS" const="True" glidein_publish="False" job_publish="False" parameter="True"
publish="False" type="string" value="default"/>
      <attr name="GLEXEC_BIN" const="True" glidein_publish="False" job_publish="False" parameter="True"
publish="True" type="string" value="OSG"/>
      <attr name="GLIDEIN_CMSSite" const="True" glidein_publish="True" job_publish="True"
parameter="True" publish="True" type="string" value="T2_US_UCSD"/>
      <attr name="GLIDEIN_Max_Walltime" const="True" glidein_publish="False" job_publish="False"
parameter="True" publish="True" type="int" value="171000"/>
      <attr name="GLIDEIN_SEs" const="True" glidein_publish="True" job_publish="True" parameter="True"
publish="True" type="string" value="bsrm-1.t2.ucsd.edu"/>
      <attr name="GLIDEIN_Site" const="True" glidein_publish="True" job_publish="True" parameter="True"
publish="True" type="string" value="UCSD"/>
      <attr name="GLIDEIN_Supported_VOs" const="True" glidein_publish="False" job_publish="False"
parameter="True" publish="True" type="string"
value="CMS,GLOW,GPN,HCC,NEBioGrid,GLUEX,UCSDRok,NWICG,glowVO,HCCLONG,CMST2UCSD,EngageVO"/>
      <attr name="USE_CCB" const="True" glidein_publish="True" job_publish="False" parameter="True"
publish="True" type="string" value="True"/>
   </attrs>
   ...
```

Generated Submit File

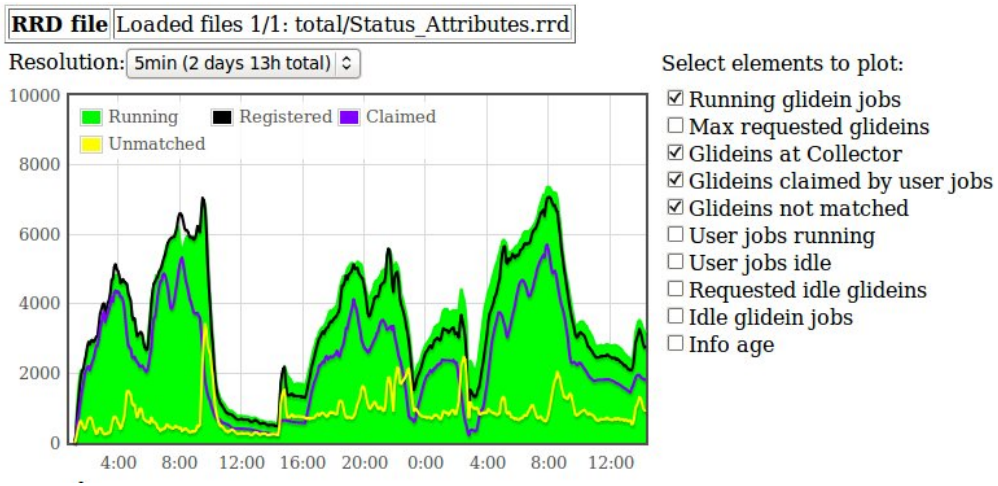
```
# File: job.condor
#
Universe = grid
Grid_Resource = gt2 osg-gw-2.t2.ucsd.edu:2119/jobmanager-condor
globus_rsl = (queue=cms)(jobtype=single)
Executable = glidein_startup.sh
copy_to_spool = True
Arguments = -v $ENV(GLIDEIN_VERBOSITY) -cluster $(Cluster) -name Production_v4_0 -entry
CMS_T2_US_UCSD_gw2 -clientname $ENV(GLIDEIN_CLIENT) -subcluster $(Process) -schedd $ENV(GLIDEIN_SCHEDD)
 -factory UCSD -web http://glidein-1.t2.ucsd.edu:8319/glidefactory/stage/glidein_Production_v4_0 -sign
11e2b24555b0023117c92ed1388f68d2dc635786 -signentry 8a66bb4fa6b94ac1aab7c60258b6b561328391cb -signtype
sha1 -descript description.b48i2k.cfg -descriptentry description.b47eLI.cfg -dir Condor
-param_GLIDEIN_Client $ENV(GLIDEIN_CLIENT) $ENV(GLIDEIN_PARAMS)
+GlideinFactory = "UCSD"
+GlideinName = "Production_v4_0"
+GlideinEntryName = "CMS_T2_US_UCSD_gw2"
+GlideinClient = "$ENV(GLIDEIN_CLIENT)"
+GlideinX509Identifier = "$ENV(GLIDEIN_X509_ID)"
+GlideinX509SecurityClass = "$ENV(GLIDEIN_X509_SEC_CLASS)"
+GlideinWebBase = "http://glidein-1.t2.ucsd.edu:8319/glidefactory/stage/glidein_Production_v4_0"
+GlideinLogNr = "$ENV(GLIDEIN_LOGNR)"
+GlideinWorkDir = "Condor"
Transfer_Executable = True
transfer_Input_files =
transfer_Output_files =
WhenToTransferOutput  = ON_EXIT
Notification = Never
+Owner = undefined
Log =
/var/gfactory/clientlogs/user_$ENV(GLIDEIN_USER)/glidein_Production_v4_0/entry_CMS_T2_US_UCSD_gw2/condo
r_activity_$ENV(GLIDEIN_LOGNR)_$ENV(GLIDEIN_CLIENT).log
Output =
/var/gfactory/clientlogs/user_$ENV(GLIDEIN_USER)/glidein_Production_v4_0/entry_CMS_T2_US_UCSD_gw2/job.$
(Cluster).$(Process).out
Error =
/var/gfactory/clientlogs/user_$ENV(GLIDEIN_USER)/glidein_Production_v4_0/entry_CMS_T2_US_UCSD_gw2/job.$
(Cluster).$(Process).err
stream_output = False
stream_error  = False
Queue $ENV(GLIDEIN_COUNT)
```

Validation

- We do basic validation on sites
- Before a glidein starts it tests the WN environment to ensure it can run
- Frontends can include their own validation scripts to further ensure they have everything on the WN they need
  - If a validation script returns with a non-zero value the glidein terminates and reports validation error
  - This prevents glideins from starting user jobs if validation isn't passed first
- Validation errors are tracked in the monitoring making it easier to find and troubleshoot failing glideins

Site Debugging

- We have a set of monitoring tools to ensure glideins are running as expected
- If we see something is wrong we first check if it can be fixed from our end
  - Else we collect any useful debugging info from the logs / monitoring and
  - Open service tickets and work closely with the site to debug



XML last update: Mon Mar 7 06:19:56 2011

		Stat			
Entry Name		**Running**	**Idle**	**Waiting**	**Pending**
CMS_T2_US_UCSD_gw2	↑	**404**	**50**	**0**	**50**
CMS_T2_US_UCSD_gw4	↑	**344**	**66**	**0**	**66**
CMS_T2_US_Nebraska_Husker	↑	**117**	**93**	**0**	**93**
HCCHTPC_T2_US_Purdue_Lepton	↑	**84**	**2**	**0**	**2**
CMS_T2_US_Nebraska_Red	↑	**91**	**73**	**0**	**73**

Types of problems

- Site down for maintenance
- Stuck idle glideins
- Glideins in held state
- Grid authentication failures
- Broken/full disks
- Missing/corrupted WN software
- Network problems
- Glidein authentication problems

Example – Condor-G error

- Check Factory Status monitoring page
  - High number of Held jobs at TAMU

Entry Name		Name	Running	Idle	Waiting	Pending	Staging in	Staging out	Unknown	Held
CMS_T3_US_TAMU	↑	Now	0	25	25	0	0	0	0	1000
		2 hours	0	20.54	20.54	0	0	0	0	1001.2
		24 hours	0	19.44	19.44	0	0	0	0	1001.75

- Hold reason shows authentication error

```
HoldReason = "Globus error 7: authentication with the remote server failed"
```

- We open a ticket with the Site, providing the DN used
  - Not much more we can do

Example – Condor-G error

- Check Factory Status monitoring page
  - High number of Held jobs at IIHE

Entry Name		Name	Running	Idle	Waiting	Pending	Staging in	Staging out	Unknown	Held
CMS_T2_BE_IIHE_cream01_cms		2 hours	150.31	0	0	0	0	0	0.01	0
		24 hours	59.28	0.7	0	0.7	0	0	1.87	0
		7 days	60	4.59	0.02	4.57	0	0	0.67	231.84

- Hold reason shows timeout error

```
HoldReason = "CREAM error: CREAM_Set_Lease Error: Received NULL fault; the error is due to another
cause: FaultString=[connection error] - FaultCode=[SOAP-ENV:Client] - FaultSubCode=[SOAP-ENV:Client] -
FaultDetail=[Connection timed out]"
```

- IIHE not advertising downtime, so opening ticket
- IIHE comes back claiming they are up and running, and other users are happily using their resources
- After more debugging, turns out we have been blacklisted

Example – Network error

- Check our daily analyze_entries email
  - 100% validation errors at UC Riverside site

```
Per Entry (all frontends) stats for the past 24 hours.

                                  strt fval 0job |  val idle  wst badp |  waste    time   total
HCC_BR_UNESP                        0%   0%   6% |   0%  12%  12%  12% |   3668   28467  | 9147
CMS_T2_US_UCSD_gw4                  2%   0%  32% |   0%  12%  12%  13% |   2998   23969  | 7959
...
CMS_T3_US_UCR_top                 100% 100% 100% | 100%   0% 100% 100% |    135     135  |  405
```

- Search glidein error logs for problem

```
Tue Dec 21 18:39:09 PST 2010 Failed to load file 'description.acgcUc.cfg' from 'http://glidein-
1.t2.ucsd.edu:8319/glidefactory//stage/glidein_Production_v3_1' using proxy 'charm.hep.int:3128'
```

- First verify nothing is wrong with our webserver
  - Everything looked fine on our end so we opened a service ticket at UCR
- UCR confirmed that their squid was down and restarted it

Example – Network error

- Check our daily analyze_queues email
  - 91% of SBGrid glideins don't register with the collector

```
Frontend stats for the past 24 hours, units = Slots.

              Run    Held    Idle  Unknwn | Pending    Wait   StgIn  StgOut |  RunDiff IdleDiff   %RD
CMS         11.2K    1.1K    1.7K     2.7 |    1.7K    41.4     0.1    82.9 |    166.9    591.6    1%
...
SBGrid      415.1     0.1   368.7     0.0 |   368.6     0.0     0.0    31.6 |   -379.2    296.6  -91%
```

- But no obvious errors in the glidein logs!
  - Glideins just don't show up in the collector
  - Only glideins working are those at Harvard
- Previous experience tells us this could be a firewall issue
  - Although Harvard network admins claim it cannot be
- We arrange for a network test between UCSD and Harvard
  - Prove UDP traffic (but not TCP) is indeed being filtered!

Example – WN problems

- Check our daily analyze_entries email
  - 16% errors at Florida Tech

```
Per Entry (all frontends) stats for the past 24 hours.

                                          strt fval 0job |  val idle  wst badp |  waste    time   total
CMS_T2_PT_LIP_Lisbon_ce02_cmsgrid_x86_64   0%   0%  38% |   2%  68%  73%  86% |     47      63  |  105
CMS_T3_US_FIT_uscms1                      16%  11%  55% |  12%  64%  78%  88% |     44      56  |  117
NEBIO_US_Harvard_HMS_East                  0%   0%  25% |   0%  17%  19%  19% |     44     232  |   81
```

- Search glidein error logs for problem

```
/mnt/nas0/OSG/GRID/setup.sh: line 208: /nas0/OSG/GRID/vdt/etc/vdt-globus-options.sh: No such file or
directory
Mon Mar 21 11:05:26 EDT 2011 GLOBUS_PATH not defined and /nas0/OSG/GRID/globus/etc/globus-user-env.sh
does not exist.
```

- While a large number of jobs failed, they were restricted to a single node
  - We provide this info to the site
  - Site discovers it was due to a bad reinstall of WN software
- We routinely catch when specific nodes fail on a site, often before the site notices

Example – WN problems

- Check our daily analyze_entries email
  - 25% errors in Rome

```
Per Entry (all frontends) stats for the past 24 hours.

                                          strt fval 0job |  val idle  wst badp |  waste    time   total
CMS_T2_PT_LIP_Lisbon_ce02_cmsgrid_x86_64   0%   0%  38% |   2%  68%  73%  86% |     47      63  |  105
CMS_T2_US_Roma1                           25%  25%  55% |  20%  64%  78%  88% |     76      55  |  222
NEBIO_US_Harvard_HMS_East                  0%   0%  25% |   0%  17%  19%  19% |     44     232  |   81
```

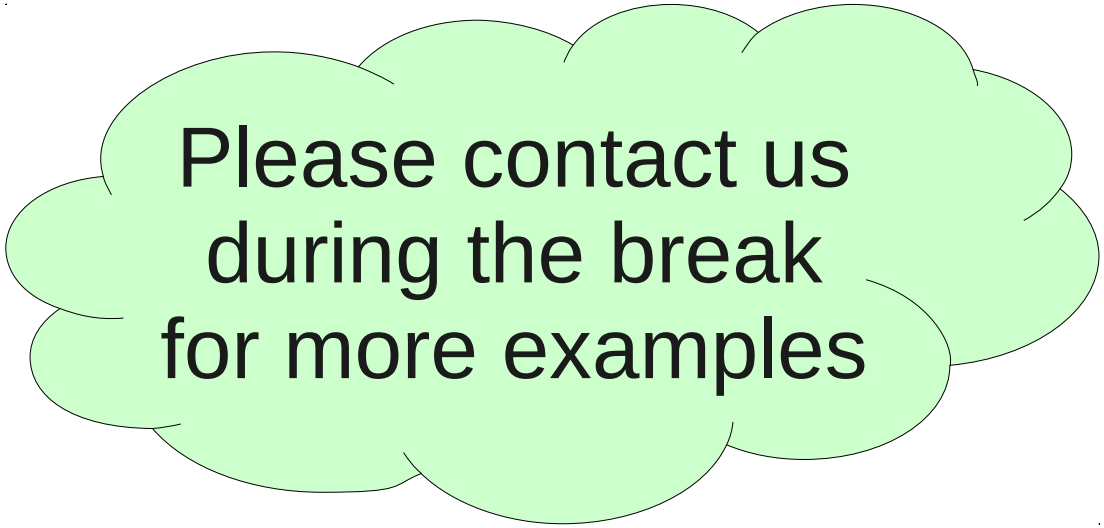
- Search glidein error logs for problem

```
cmsset_default.sh not found!
Looked in /cmsset_default.sh
and /cmssoft/cms/cmsset_default.sh
=== Validation error in
/home/cms058/globus-tmp.cmsrm-wn070.23125.0/glide_f23329/client/discover_CMSSW.sh
===
```

- This time it is a VO provided script that is failing (again just on a few nodes)
  - Must first contact the VO about what is being tested
  - Then open a ticket with site

Disclaimer

- The examples shown are just a tiny fraction of those we discover and fix
- Time is limited, so we selected just a few that could be fit on slides



Summary

- The Grid is an error-prone place to live in
  - Not surprising, given the size
- Exposing users directly to it can be expensive
  - Wasted time debugging the infrastructure
  - Users not using it due to bad experience
- Hiding the Grid from the users helps
  - But someone still needs to do the dirty job
- The glideinWMS approach concentrates this in the hands of only a few people
  - Experienced, expert → more efficient
  - Economies of scale lower the TCO

Acknowledgements

- Many thanks to the glideinWMS and Condor teams for providing the great software
- Many thanks to OSG for their continuing support
- This work is partially sponsored by
  - the US Department of Energy under Grant No. DE-FC02-06ER41436 subcontract No. 647F290 (OSG), and
  - the US National Science Foundation under Grants No. PHY-0612805 (CMS Maintenance & Operations), and OCI-0943725 (STCI).

Copyright notice

- Several images in this presentation are copyright of ToonADay.com
  and have been licensed by
  Igor Sfiligoi for use in his presentations
- Any other use strictly prohibited