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

- 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/>

  <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>

  ...
</entry>
# 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 $ENV(GLIDEIN_SCHEDD)
+Factory UCSD -web http://glidein-1.t2.ucsd.edu:8319/glidefactory/entry/condor
+GlideinFactory = "UCSD"
+GlideinEntryName = "$ENV(GLIDEIN_ENTRY_NAME)"
+GlideinClient = "$ENV(GLIDEIN_CLIENT)"
+GlideinX509Identifier = "$ENV(GLIDEIN_X509_ID)"
+GlideinX509SecurityClass = "$ENV(GLIDEIN_X509_SECURITY_CLASS)"
+GlideinWebBase = "http://glidein-1.t2.ucsd.edu:8319/glidefactory/entry/condor"
+GlideinLogNr = "$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
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

<table>
<thead>
<tr>
<th>Entry Name</th>
<th>Name</th>
<th>Running</th>
<th>Idle</th>
<th>Waiting</th>
<th>Pending</th>
<th>Staging in</th>
<th>Staging out</th>
<th>Unknown</th>
<th>Held</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Now</td>
<td>0</td>
<td>25</td>
<td>25</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1000</td>
</tr>
<tr>
<td>CMS_T3_US_TAMU</td>
<td>2 hours</td>
<td>0</td>
<td>20.54</td>
<td>20.54</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1001.2</td>
</tr>
<tr>
<td></td>
<td>24 hours</td>
<td>0</td>
<td>19.44</td>
<td>19.44</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1001.75</td>
</tr>
</tbody>
</table>

- 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

<table>
<thead>
<tr>
<th>Entry Name</th>
<th>Name</th>
<th>RunningIdle</th>
<th>Waiting Pending</th>
<th>Staging In</th>
<th>Staging out</th>
<th>UnknownField</th>
</tr>
</thead>
<tbody>
<tr>
<td>CMS_T2_BE_IIHE_cream01_cms</td>
<td>2 hours</td>
<td>150.31</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>24 hours</td>
<td>59.28</td>
<td>0.7</td>
<td>0.7</td>
<td>0</td>
<td>1.87</td>
</tr>
<tr>
<td></td>
<td>7 days</td>
<td>60</td>
<td>4.59</td>
<td>0.02</td>
<td>4.57</td>
<td>0</td>
</tr>
</tbody>
</table>

- Hold reason shows timeout error

```
HoldReason = "CREAM error: CREAM_Set_Lease Error: Received NULL fault; the error is due to another
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

<table>
<thead>
<tr>
<th>Per Entry (all frontends) stats for the past 24 hours.</th>
<th>strt fval 0job</th>
<th>val idle wst badp</th>
<th>waste</th>
<th>time</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>HCC_BR_UNESP</td>
<td>0% 0% 6%</td>
<td>0% 12% 12% 12%</td>
<td>3668</td>
<td>28467</td>
<td>9147</td>
</tr>
<tr>
<td>CMS_T2_US_UCSD_gw4</td>
<td>2% 0% 32%</td>
<td>0% 12% 12% 13%</td>
<td>2998</td>
<td>23969</td>
<td>7959</td>
</tr>
<tr>
<td>CMS_T3_US_UCR_top</td>
<td>100% 100% 100%</td>
<td>100% 0% 100% 100%</td>
<td>135</td>
<td>135</td>
<td>405</td>
</tr>
</tbody>
</table>

- 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

<table>
<thead>
<tr>
<th>Run</th>
<th>Held</th>
<th>Idle</th>
<th>Unknwn</th>
<th>Pending</th>
<th>Wait</th>
<th>StgIn</th>
<th>StgOut</th>
<th>RunDiff</th>
<th>IdleDiff</th>
<th>%RD</th>
</tr>
</thead>
<tbody>
<tr>
<td>CMS</td>
<td>11.2K</td>
<td>1.1K</td>
<td>1.7K</td>
<td>2.7</td>
<td>1.7K</td>
<td>41.4</td>
<td>0.1</td>
<td>82.9</td>
<td>166.9</td>
<td>591.6</td>
</tr>
<tr>
<td>SBGrid</td>
<td>415.1</td>
<td>0.1</td>
<td>368.7</td>
<td>0.0</td>
<td>368.6</td>
<td>0.0</td>
<td>0.0</td>
<td>31.6</td>
<td>-379.2</td>
<td>296.6</td>
</tr>
</tbody>
</table>

- 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

<table>
<thead>
<tr>
<th>Per Entry (all frontends) stats for the past 24 hours.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Start fval 0job</td>
</tr>
<tr>
<td>----------------</td>
</tr>
<tr>
<td>CMS_T2_PT_LIP_Lisbon_ce02_cmsgrid_x86_64</td>
</tr>
<tr>
<td>CMS_T3_US_FIT_uscsm1</td>
</tr>
<tr>
<td>NEBIO_US_Harvard_HMS_East</td>
</tr>
</tbody>
</table>

- 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

<table>
<thead>
<tr>
<th>Per Entry (all frontends) stats for the past 24 hours.</th>
</tr>
</thead>
<tbody>
<tr>
<td>strt fval 0job</td>
</tr>
<tr>
<td>----------------</td>
</tr>
<tr>
<td>CMS_T2_PT_LIP_Lisbon_ce02_cmsgrid_x86_64</td>
</tr>
<tr>
<td>CMS_T2_US_Roma1</td>
</tr>
<tr>
<td>NEBIO_US_Harvard_HMS_East</td>
</tr>
</tbody>
</table>

- 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.

Please contact us during the break for more examples.
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