CMS Glidein Factory - Bootstrapping a Condor Pool Spanning Computing Sites Around the Globe

Amjad Kotobi
On behalf of the CMS Submission Infrastructure Group and
OSG Factory Operations Team
June 2017
CMS Submission Infrastructure

Global Pool

There are two different types of jobs in pool:
1. CMS Remote Analysis Builder (CRAB3)
2. Central production (WMAgent)

Global pool was created in 2014 for flexibility to use entire CMS resources for different kind of workflows.

Submission infrastructure is pilot based with two main components:
1. HTCondor Pool
2. GlideinWMS (Glidein based workload management system, e.g Frontend, Factory)
CMS Submission Infrastructure

GlideinWMS Factory

GlideinWMS Frontend

Schedd

CRAB

WMAgent

Central Manager

Collector

Negotiator

Site A

Startd

Site B

Startd

Site C

Startd

Site D

Startd

Cloud

Startd
GlideinWMS Components

- **Glidein**: starts condor startd on the grid site.

- **Frontend**: Polls user jobs and make sure there are enough glideins for users job and make resource request to glidein factory.

- **Factory**: Receives request from frontend to submit glideins to grid sites.

- **WMS Collector**: A condor collector, keeps factory entry ClassAds and frontend request ClassAds.
First Stage Matchmaking

- Factory advertises entries ClassAds to the WMS collector.
- Factory has each entry description which advertises so glidein able to land on particular grid resources.
- Frontend check WMS collector to exert match expression against entry and user jobs.
- Match expression comes from frontend config and VO decides to where send user jobs to run.
- If there is no glidein so frontend will make request for more submission.
GlideinWMS Factory Role

1. Advertise itself
2. Listen to frontend request
3. Submit glideins

Glidein Factory Role:

Job Pressure

Resource Request Pressure

GlideinWMS Factory

Glidein Submission

Grid Site Resources
What is GlideinWMS Factory

- Glidein factory actually is a workload management system (WMS), uses pilot submission model to send jobs to grid resources.

- GlideinWMS factory works on top of HTCondor and heavily dependent on it, factory plays as schedds for pilot jobs.

- Glidens are actually pilot only glideinWMS calls it pilot.

- Pilot is an actual grid job and HTCondor job by itself that after reach to grid site calls real user jobs, in other words glideins are placeholder on remote resources.
Schematic View of GlideinWMS Factory
Schematic View of GlideinWMS Factory

Collector

Frontend Req

Factory ClassAds

GlideinWMS Factory

Grid Sites

Site A

Gate Keeper

Site B

Gate Keeper

Glidein

Schedd
Schematic View of GlideinWMS Factory
Schematic View of GlideinWMS Factory
Schematic View of GlideinWMS Factory

Collector
- Frontend Req
- Factory ClassAds

GlideinWMS Factory
- Glidein
- Schedd

Grid Sites
- Site A
  - Gate Keeper
- Site B
  - Gate Keeper
  - Glidein Pilot

1. Collector
2. Frontend Req
3. Factory ClassAds
4. GlideinWMS Factory
5. Schedd
6. Grid Sites
Schematic View of GlideinWMS Factory

Collector

Frontend Req

Factory ClassAds

GlideinWMS Factory

Glidein

Schedd

Grid Sites

Site A

Gate Keeper

Site B

Gate Keeper

HTCondor

Startd
Second Stage Matchmaking

- **After** pilots/glideins start running on grid sites resources
- **Then** launched startds get connected back to the global pool HTCondor
- **Finally**, the negotiator matches job requests to the advertised available resources.
Importance of Factory to CMS

• Glideins run validation scripts to make sure environment suits users job

• Users job never starts in broken worker node and will find another match glideins

• Glideins play a role of placeholder to reserve resources for users job

• Glideins do not tie to a single user job during its lifetime

• Factory holds grid sites configurations and less trouble on site admin side.

• Factory can serve several frontends e.g ITB and Global

• Submitting jobs get simpler and resource shows as condor pool
CMS Support Factories

GlideinWMS Factory

FNAL

GlideinWMS Factory

UCSD

GlideinWMS Factory

OSGGOC

CERN CMS

GlideinWMS Factory
GlideinWMS Factory

• Each factory has CMS global frontend to get resource request pressure

• Each factory has **nine** schedds

• Factory has collector

• Factory can send single and multi-core glideins to grid site

• Redundancy of factories

• Factory submit jobs to grid sites by **Condor-G**

• Each factory can support many frontends VOs
Factory Submission to Non-Traditional Resources

- Factory covers glideins submission interface to non-traditional resource e.g. Tier-0 cloud and NERSC

- CERN agile infrastructure (CERN AI) is an Openstack CERN cloud for resources.
- HTCondor and GlideinWMS used for job execution backbone.
- EC2/GlideinWMS interface make it feasible the pilot itself starts the VM and integrated in the image.
# Example Factory Entry of Grid Site

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Value</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>FactoryType</td>
<td>production</td>
<td></td>
</tr>
<tr>
<td>GCB_ORDER</td>
<td>NONE</td>
<td></td>
</tr>
<tr>
<td>GLEXEC_BIN</td>
<td>gLite</td>
<td></td>
</tr>
<tr>
<td>GLIDEIN_CMSSite</td>
<td>T1_DE_KIT</td>
<td></td>
</tr>
<tr>
<td>GLIDEIN_CPUS</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>GLIDEIN_Country</td>
<td>DE</td>
<td></td>
</tr>
<tr>
<td>GLIDEIN_Gatekeeper</td>
<td>arc-6-kit.gridka.de</td>
<td></td>
</tr>
<tr>
<td>GLIDEIN_GlobusRSL</td>
<td>(queue=grid)(count=8)(memory=2500)(runtimeenvironment=ENV/GLITE)</td>
<td>GlobusRSL</td>
</tr>
<tr>
<td>GLIDEIN_GridType</td>
<td>nordugrid</td>
<td>GridType</td>
</tr>
<tr>
<td>GLIDEIN_MaxMemMBs</td>
<td>20240</td>
<td></td>
</tr>
<tr>
<td>GLIDEIN_Max_Waittime</td>
<td>216000</td>
<td></td>
</tr>
<tr>
<td>GLIDEIN_REQUIRED_OS</td>
<td>rhel6</td>
<td></td>
</tr>
<tr>
<td>GLIDEIN_REQUIRE_GLEXEC_USE</td>
<td>False</td>
<td></td>
</tr>
<tr>
<td>GLIDEIN_REQUIRE_VOMS</td>
<td>False</td>
<td></td>
</tr>
<tr>
<td>GLIDEIN_Req_MUPI_gLExec</td>
<td>False</td>
<td></td>
</tr>
<tr>
<td>GLIDEIN_ResourceName</td>
<td>FZK-LCG2</td>
<td></td>
</tr>
<tr>
<td>GLIDEIN_RelTime</td>
<td>108000</td>
<td></td>
</tr>
<tr>
<td>GLIDEIN_RelTime_Spread</td>
<td>7200</td>
<td></td>
</tr>
<tr>
<td>GLIDEIN_SEs</td>
<td>cmsrm-kit.gridka.de</td>
<td>RemoveSleep</td>
</tr>
<tr>
<td>GLIDEIN_Site</td>
<td>KIT</td>
<td>RequireGlideInGlexecUse</td>
</tr>
<tr>
<td>GLIDEIN_SlotsLayout</td>
<td>fixed</td>
<td>RequireVomsProxy</td>
</tr>
<tr>
<td>GLIDEIN_SupportedAuthenticationMethod</td>
<td>grid_proxy</td>
<td>StartupDir</td>
</tr>
<tr>
<td>GLIDEIN_Supported_VOs</td>
<td>CMS</td>
<td>SubmitCluster</td>
</tr>
<tr>
<td>GLIDEIN_TrustDomain</td>
<td>grid</td>
<td>SubmitSleep</td>
</tr>
<tr>
<td>GLIDEIN_Verbosity</td>
<td>std</td>
<td>SubmitSleep</td>
</tr>
<tr>
<td>GLIDEIN_WorkDir</td>
<td>TMPDIR</td>
<td>TrustDomain</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Grid Site Entries

Grid sites may have many entries in factory with different gatekeepers or queues that connected to diverse resources

<table>
<thead>
<tr>
<th>CMSHTPC_T1_DE_KIT_arc-1</th>
<th>↑</th>
<th>6</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>CMSHTPC_T1_DE_KIT_arc-2</td>
<td>↑</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>CMSHTPC_T1_DE_KIT_arc-3</td>
<td>↑</td>
<td>0</td>
<td>9</td>
<td>9</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>CMSHTPC_T1_DE_KIT_arc-4</td>
<td>↑</td>
<td>0</td>
<td>4</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>CMSHTPC_T1_DE_KIT_arc-5</td>
<td>↑</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>CMSHTPC_T1_DE_KIT_arc-6</td>
<td>↑</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
CMS Support Factory Usage
CMS Support Factory Usage

<table>
<thead>
<tr>
<th>Date</th>
<th>Usage</th>
</tr>
</thead>
<tbody>
<tr>
<td>May 6</td>
<td>20000</td>
</tr>
<tr>
<td>May 9</td>
<td>15000</td>
</tr>
<tr>
<td>May 12</td>
<td>10000</td>
</tr>
<tr>
<td>May 15</td>
<td>15000</td>
</tr>
<tr>
<td>May 18</td>
<td>10000</td>
</tr>
<tr>
<td>May 21</td>
<td>15000</td>
</tr>
<tr>
<td>May 24</td>
<td>10000</td>
</tr>
<tr>
<td>May 27</td>
<td>15000</td>
</tr>
<tr>
<td>May 30</td>
<td>10000</td>
</tr>
<tr>
<td>Jun 2</td>
<td>15000</td>
</tr>
</tbody>
</table>

**Running**

CERN

OSGGOC
CMS Support Factory Usage
CMS Support Factory Usage

Running

May 6  May 9  May 12  May 15  May 18  May 21  May 24  May 27  May 30  Jun 2
Acknowledgements

• Special thanks to **HTCondor** developers for their support and suggestions

• **Factory Operations team**
  - Jeff Dost, Krista Larson, Marian Zvada, Marty Kandes

• **Submission infrastructure team leaders**
  - James Letts, Antonio Perez-Calero Yzquierdo, David Mason

• **GlideinWMS developers**

• **CMS Global Pool**
  - Diego Davila

• **Brian Bockelman, Farrukh Aftab Khan**