**CMS feedback to OSG WBS for year 5**

The purpose of this document is for US CMS to provide concrete feedback to OSG with regard to expectations from OSG in year 5. We divide this up into ongoing operations support, and a prioritized list of work items, and a timeline when we need them.

**Note:**

This is a living document. It will be reviewed roughly once every 3 months throughout year 5 to indicate what has been accomplished, missed, and record changes. We identify as accomplished by marking things in green, and missed in red. Red things turn green as they are completed.

We note additions by prefacing them with a date.

We note things that get dropped by prefacing them with a date, and changing to gray.

**Ongoing Operations:**

There’s a long list of operations services that OSG provides, and CMS benefits from. These include:

* Security context. This includes operational, policy, software, validation.
* BDII at the GOC.
* GOC ticketing system.
* Accounting and RSV, and related reporting to WLCG
* Software cache and software lifecycle management via the VDT, including the OSG cache, as well as deployment testing, incl. vtb and itb.
* Variety of validation, including VOMS and GIP.
* Registration and administrative services
* Support, including OSG-storage (office hour, ticketing), and CE deployment.
* glideinWMS operations support

Among all of these, the bdii is special in that the complete path from GIP -> CEMON -> bdii -> wLCG is crucial for CMS operations. If the bdii at the GOC disappears then all of OSG disappears from view for the CMS job submission infrastructure.

As a result of this special dependency of CMS on this service, we request that OSG plan ahead, and develop mechanisms in time to increase the [BDII] scale as the load increases. <Quick>

Almost as crucial is the glideinWMS operations support. This is new in year 5 and we will elaborate more on this below.

**Prioritized Work Items:**

Priority of an item is indicated by a number in () after the item. The lower the number the higher the priority.

**WLCG Issues:**

Throughout year 5:

* Manage Change. We expect changes to occur throughout the year. We expect OSG to work with WLCG, and be pro-active about upcoming changes. We propose as success metric the continued availability of accounting and site availability/reliability information <Fraser, Quick> in all monthly reports of WLCG throughout the year. At present, we expect the possible introduction of space accounting, and capacity monitoring <Bockelman>(CPU and storage). (1)
* Work with WLCG to ensure continued interoperation between OSG and EGEE sites. In particular, security team provides feedback over security policies that are produced by JSPG (WLCG's policy organ). The feedback will alert to any policy items that can break interoperability between OSG-EGEE, and items that cannot reasonably be achieved by OSG or CMS (1).

**Site Operations Improvements:**

January 2010:

* Improvements to CRL updateing reliability. (fkw is not sure if this was completed, or if what is installed at UCSD is a local hack) (2)

Thoughout year 5:

* (1) The cost of operating the middleware, especially storage is still too high. We need lower cost solutions for both T3 deployment. <Levshina, Fraser> Lowering cost implies:
  + Less complex middleware deployment and operations
  + More robustness, i.e. reliability over time that leads to decreased need for operator intervention.
  + Maybe more and better training, especially with regard to operations. Maybe this could be done in combination with experienced site admins. There are a lot of “gremlins” deployed at sites like FNAL and BNL in order to guarantee successful operations, especially of storage.
* We have a “downtime goal” of 6-12 hours for an SE upgrade <Levshina>. This is pure downtime, and does not include “drain-off”. (2)
* We worry about transitioning away from GT2. We depend on OSG to provide advice. This includes documentation on the relative performance characteristics of GT2 vs GRAM5 or other solutions. It is not clear to us on what timescale this will happen. CMS is reactive here, rather than driving the issue. (1)

**Storage Issues:**

Throughout year 5:

* CMS presently depends on dCache, hdfs, BeStMan, globus gridftp, and we will be adding Xrootd use in year 5. We thus require continued packaging, testing, understanding of scaling issues for all of these. Some of this may be accomplished via satellite projects like ANI, or others. However, we would like OSG to provide the coordination role that brings all of this together. In the following we list some of the details.

October 2010:

* We have reached sustained rates of ~100Hz for lcg-ls for BeStMan. BeStMan2 was released, and is expected to become the default within the next few months. It is supposed to fix a known overload behaviour where the service fails, requiring restart, if a certain scale is exceeded.

We need verification that BeStMan2 indeed fixes the known problem, and

does not introduce other problems. (2)

We need the scaling behaviour of BestMan2 verified. Initial tests show

that the scaling behaviour is very different between BetsMan and

BestMan2. These tests need to be concluded, and documented. (1)

At present, the deployment advice for BeStMan2 is still in flux. This needs

to conclude. (1)

* hadoop support was added to the VDT in year 5. This needs to be maintained. In particular, we need support for the next hdfs release. (2)

January 2011:

* We need support for Xrootd as WAN access protocol in the context of diskless T3. (2)

**WMS Issues:**

Throughout Year 5:

* CMS data analysis depends on a joint glideinWMS installation with OSG. At present, three VOs other than CMS use the jointly operated service. Within the last year, the operations effort came primarily through CMS as it was not clear that there is interest by other VOs in this service. We believe such interest is clearly established now, and we’d like to see increased OSG effort to support operations. (1)
* CMS will start operations of small scale parallel applications. This will initially be restricted in production to the Tier-1 sites. It would be good for OSG to stay abreast of this change as it will become standard for CMS in a year or two. Unless there is a sensible way to schedule jobs opportunistically in this environment, OSG is likely to loose access to CMS resources once CMS starts scheduling full nodes instead of cores at its T2 centers. OSG could benefit from understanding the performance benefits that drive CMS towards Multi-Core applications. (3) Some of the concrete issues to resolve are:
  + The condor gratia probe does not report how many cores were used by the job. As most of the US-CMS sites use condor, this is a serious issue.
  + There is a need to improve the accounting of memory usage by multi-process jobs. Just summing up the RSS or VSIZE isn’t very useful
* CMS would like to have the capability to transparently use Cloud resources, especially Magellan, and Grid resources. As the paradigm within which CMS will be doing this is glideinWMS, it would be good for OSG to play a role in making this happen. Whatever can be learned from it for CMS will be immediately applicable to other OSG customers via glideinWMS. (3)
* As more VOs in CMS depend on glideinWMS, we need to understand change control better. OSG has a role to play here. (1)
* OSG should start thinking more about privilege separation for glideins and proper accounting. (3)

**Other Issues:**

Throughout year 5:

* We expect significant ramp-up of T3s in CMS, and expect support in form of training and Q&A, <Fraser> as well as trouble shooting of deployments. This is an issue for both CE and SE. (1)
* We would like to see more effort put into native packaging. This is becoming more important as we take a hard look at total cost of ownership. It ought to also have become easier as a significant portion of our core software has become of part of packaged Linux distributions already. As we only support Linux platforms, it seems silly not to benefit from this trend for CMS.