Effort Summary for Production Gratia Collectors Upgrade to Gratia 1.08

January 5, 2012

Steve Timm, Dan Yocum, Keith Chadwick

Introduction:

This document is a short description of the steps that were taken to develop, test, and execute a safe upgrade procedure to Gratia 1.08, and identify the steps that caused the most extra unplanned effort.

Timeline:

Aug. 2010 - Gratia 1.08 is promised.

Jan. 2011 - gratia-osg-transfer database gets so far behind that we cannot run housekeeping without falling behind, we disable housekeeping and from then on the transfer database is growing without bounds.

Jul. 14, 2011 - Gratia 1.08.00 is released, 11 months behind schedule.

Aug. 15, 2011 - Dan Yocum begins tests with gratia-osg-transfer database. These are done by restoring copy of database from backups in FermiCloud. First try fails due issues with ZRM unzipping into a too-small partition (30GB uncompressed turns into 320GB sql dump file). Second through fifth (?!) attempts failed due to issues with pdflush, kjournald, huge table memory support (HugeTLB), O_DIRECT, virtio, and KVM. Two additional attempts to restore the database failed due to site-wide power outages. Finally successfully restored a copy of the database on Sep. 27, 2011.

Sep. 12, 2011 - System gratia12 (production server) hits a motherboard fault, fixed the next day.

Sep. 30, 2011 - Attempts to do a collector upgrade/ database schema update on the transfer database fail due to more unanticipated foreign key constraints.

Oct. 5, 2011 - Request to the gratia developers for documentation on parts of the missing installation procedure.

Oct. 6, 2011 - A procedure of pruning table by table begins, finally converges.

Oct. 7, 2011 - Begin trying to do the same to the gratia-osg-prod database. Hit more and different ForeignKey constraint issues.

Oct 11, 2011 - This process converges several iterations later.

Oct 18, 2011 - We then find that the gratia reporter doesn't start up correctly. Attempt and eventually succeed to get a bit of effort from P. Constanta, root cause there is a missing MySQL user in the database who is or should be the owner of the trigger.

Oct 27, 2011 - There was so much data that the nightly backups of the database started taking more

than 24 hours, put over the top in part by a huge influx of junk data from AGLT2. We decided to accelerate the pruning of the gratia-osg-transfer database and get that done before the upgrade. Unexpected glitches in this procedure made it take about 8 hours rather than the 2 that we expected. That got backups back to a reasonable length.

Nov. 4 2011 - The reporter upgrade email discussion finally converges.

Nov. 22 2011 - Change management at Fermilab approves the change request, The **fermi-itb** collector Gratia instance upgrade is completed.

Nov. 23, 2011 - Collector **fermi-qcd** upgrade is completed.

Nov. 28, 2011 - Collector **fermi-psacct** upgrade is completed.

Nov. 29, 2011 - Collector **fermi-transfer** upgrade is completed.

Nov. 30, 2011 - Collector **fermi-osg** upgrade is completed.

At this point we realize that given the amount of data in the OSG production databases, and the amount of time taken for the smaller databases, the upgrade in question might take as long as three days for OSG production and we begin discussion with R. Quick and D. Fraser about alternatives to keep the server up and minimize downtime.

Dec. 5, 2011 - We settle on an alternative of replicating all existing OSG data to fermicloud338 (a virtual machine in FermiCloud) and using that virtual machine as the temporary collector while the upgrade is ongoing.

Dec. 5, 2011 - We discuss this plan at OSG Change Management call.

Dec. 6, 2011 - We discuss this plan on OSG Production call. Receive approval in both places.

Dec. 7, 2011 - We move the FermiCloud machine in question to FCC3 (from its location in GCC) so that it can be on the correct network to serve as an interim OSG Gratia instance during the upgrade of the real OSG Gratia instances.

Dec. 9, 2011 - The reconstituted machine begins catch-up replication.

Dec. 14, 2011 - Original scheduled beginning of the **gratia-osg-prod** upgrade. Due to two catastrophic network outages at Fermilab that day, we postpone start to Dec. 15. Also we discover that although gratia replication says that the alternate collector replication is complete, the data is not quite the same, there are some number of records that don't get included in the newly replicated database, effect is maybe 0.1%. We proceed anyway because all the initial records are still in the main DB. GOC users are notified that we'll make the change.

Dec. 15, 2011 - Upgrade actually begins. Completes late Friday night Dec. 16. The actual upgrade elapsed time was ~19 hours.

Dec. 19, 2011 - We start the process of migrating the xml files collected on fermicloud338 (the temp. collector) back to gr12x0. Using the xml files instead of database-based replication due to the

inconsistencies we saw in replication on the way there.

Dec. 22, 2011 - We notice that we are not gaining on the backlog, stuck at about 160K records, investigation shows that we have a replication loop going between gr12x0 and fermicloud338.

Dec. 23, 2011 - Replication loop is broken.

Dec. 24, 2011 - Late in evening, gr12x0 finally catches up.

Dec. 27, 2011 - We are ready to switch back

Dec. 28, 2011 - We make contact with Scott Teige and actually do switch back.

Dec. 29, 2011 – The **gratia-osg-itb** upgrade begins. New foreignkey constraint issue found that did not show up in any of the previous six updates. We dump the DB and start clean and then begin the process of restoring the previous **gratia-osg-itb** records from backup.

Dec. 29, 2011: The **gratia-osg-daily** upgrade is completed, but with issues that we only discover over the New Year's weekend, several stored procedures are missing due to a warning resulting from an inaccessible directory in the /var/lib/mysql/ tree. Directory has existed since Feb 2010 and has caused no issues on previous upgrades. Directory was removed and gratia post-upgrade.sh script is successfully re-run to load stored procedures.

Dec. 31, 2011: Disk fills up on **gratia-osg-itb** collector several times from large log and history files associated with the backups, attempts by management to intervene hose the entire machine.

Jan. 3, 2012: The **gratia-osg-itb** collector is brought back online, replication continues.

Jan 9, 2012: Planned start date for the **gratia-osg-transfer** collector update.

Analysis: What caused the most amount of effort?

The developers do not test on a database that is as complicated as the real OSG one that has six years of history built in, and evidently a few in-built inconsistencies as well.   By policy we do tests like this on an integration copy of the full database and it was very good we did, otherwise we could have been stuck with the production database in an inoperable situation for weeks and would have been sorting out these issues on the production database. The process to restore the database backups was not well documented and took a lot of Dan Yocum's time to even get the database backup restored on the new machine.

This was the first significant version upgrade that any of the current operators of the Gratia service had attempted, made more complicated by the fact that it involved a complicated database schema upgrade. Many discussions with the developers were necessary every time we hit a foreign key constraint issue. FermiGrid Services had not previously encountered the foreign constrain issues, and furthermore these issues were not anticipated by the developers, thus the corresponding corrective actions/procedures were not documented.

In addition, the lateness of the release made the existing Gratia system very fragile and there were a lot of heroic measures we had to do in parallel to this to keep the system up. The release in summer 2011

of the badly-tested and poorly-configured condor Glidein probe didn't help, throwing at least 100 garbage VO's into the database and lots of unknown VO entries too. We were also fighting a very badly documented release of the email reports that were updated on August 1. There were several facility and network incidents on the Fermilab side that slowed us down as well, not to mention a couple of hardware faults on the machines.

During the upgrade of the production collector itself, we were pressured into making the alternate collector in the cloud, something that added three days of preparation time before we could start and a week of post-upgrade time to get everything replicated back. We did keep the downtime to a minimum but at the cost of much more effort from the support staff, and there were communication issues in letting all the various stakeholders know to move their database pointers and then move them back. If we had not used an alternate machine during the upgrade, the net downtime would have been approximately 19 hours. But the overall OSG effort expended would have been much less and we would have been executing a well-tested procedure.

A final comment - the design of the Gratia database schema includes a lot of triggers and thus the multi-master MySQL replication we use elsewhere in FermiGrid can't be used to provide full high availability for Gratia.