March GDB - Amsterdam

Welcome

Introduction – NIKHEF

 Security & Other Policies

Operational Security

JG: this is a change. Previously there were a small number of ROCs. Now moving to 40 NGIs – 120 people in EGI csirt.

MM: If a site is part of EGI and WLCG, sites would report to many different bodies.

JG: Risk raised, what happens if there is a serious security incident. What about coordination between bodies?

RW: A similar structure has been working for many years.

JG: More about ownership. Who owns the incident?

RW: The sites own the data and responsibility. See my role as contributing to GRID-SEC (http://cern.ch/grid-sec).

GS: Want experiments more involved – what are you looking for? We rely on X509 infrastructure.

RW: involve them in decisions for example suspending sites.

MM: WLCG security officer will not have a separate team?

RW: That is correct.

Communication with VOs. EGEE CSIRT contact the VO or WLCG security officer?

JT: What recommendations would the WLCG security officer make?

RW: Act as a bridge between sites and MB.

JT: Would you say you strongly discourage something….

RW: The role is between formal management and …

DK The balance between risk and availability is made by the MB.

MM: The WLCG officer reports to the MB and EGEE-CSIRT reports to EGI operational security officer.

CREAM

Passing job parameters (Dennis Van Dok)

Should YAIM variable have this package on or off for CREAM?

CERN has its own stuff for LSF. My opinion is that it must be configurable. So off.

JT: Have to replace YAIM variables with local ones. Could turn it off...

DvD: YAIM function will make a symbolic link if on by default.

JG: Script is part of the distribution. What are the parameters that should be in by default: memory; wall clock, CPU time, Obtainable Wall and CPU time.

DvD: and for MPI: SMPGranularity and WholeNodes

SB: Depends on where the requirement is.. at the moment max CPU. If you can pass the request... say you want 10 hrs CPU time you want to select a queue with a default time as 10hrs. Put >10hrs. With passing parameters.. even if the default is 1hr you can get 10hrs.

SB: We are talking about the glue parameters.

MS: This parameter passing was asked for when lot of heterogeneous clusters. SO batch system could send jobs with large mem requirements to other....

JT: Multi-core and not all jobs can have over default.

MS: Glue parameters expressing queue capabilities. Now basically to do back fill.

FH: If you introduce this then do not need to specify queues at site.

JG: There is still a matching as you need to get the job to the site. WMS still needs to know the capability. ... for the set of parameters...

To run job for 1 hr what should I specify

SB: IF this is deployed then the limit should be on maxobtainable not max.

IB: This is counter-intuitive.

PC: Why should the user bother.... My job has to use up to so much CPU time and memory. This is the way we want to express our jobs. Conversion behind the scheme should not be something the user has to know about.

SB: Agree, but we would have to reengineer the WMS.

IF: The JDL is generally generated for the user.

FH: How to translate the maxobtainable requirement on my site?

PC: What are the time units? Minutes of what?

JG: Shall we encourage a few more sites (2 so far) to try this implementation?

JT: Long historical discussion… if cluster has many machine types then second on one is not the same. PC wants say 10HEPSPEC seconds on a reference machine. Wants a number of cycles. Current range is factor of 5 different.

SB: Publish the HEPSPEC rating.

JG: A script exists for Torque. What about the others?

??: Working on LSF.

DvD: The requests are out. The timescale is unclear.

MS: Do we have to take account of EGEE political correctness. Does WLCG need this anymore? Less use of WMS. Experiments just want long time slots. Super publishing is not needed.

JG: As site I have to drain my batch system before work. For 48hrs….

IF: But the experiments running aggregated jobs. Pilots may be killed but they resubmit.

JG: 48hr pilots in 48hr slot then after 24hrs will have ….

MS:

IF: A vacate signal may be more benefical.

JG: What about the memory requirements.

IF: Pilots send back how much memory is available. Pilot will only send a job which matches what it sees available.

PC: Batch system can not balance then

JG: Job starts with a requirement … requirements of pilots are different then. Criticism of mechanism so far is that it does not allow better matching.

PC: Passing parameters is on corollary that there is one queue.

How different are the memory requirments for job types: MC, simulation, reconstructions….

GS: Sites have to manage it. Separate production pilots from user pilots. These tend to go on to different queues.

IF: CMS sets a maximum. If an analysis job is about the envelope then the job needs to be looked at. Same framework overhead. Guide to sites was 1.5GB per core.

PC: We do the same.

JT: Can the issue about one unit for the time gets kicked up to Maarten's task force? [Action]

ATLAS with CREAM (Graeme Stewart)

Generally good. Issue with proxy renewal at 24hrs. Gridftp overload – jobs do a direct stage-out.

JT: How does the WMS so it?

??: Set the maximum number of connections. Could move the sandbox to the CE.

Most sites run at least 2 CEs so why not run more jobs per CE.

Big T2 next year is 3000.

MS: Thought CREAM was to allow more jobs per CE.

Load on LCG CE – how much was due to transfers?

MS: It hammers the WMS directly …

GS: Pilot factories do not run a gridftp server.

Condor submission to GRAM… condor-gap server pulls the output back. With CREAM there is no condor control. One pilot factory talks to one CE.

IF: Pilots in CMS still uses WMS.

??: Should still see this with the gridftp on the WMS.

MS: But there are many WMSes able to share the load.

IF: What are the sandboxes doing?

GS: Standard in and out. It is the number of connections and forks that is the issue.

JG: Could limit number of connections

GS: Then will get lower efficiencies for the jobs as the connection limits will lead to blocking job slots.

JG: It is a change request to condor and CREAM.

MS: Did you did some job error rate comparision between CREAM and LCG-CE.

GS: Did some tweaks… got up to 35000 jobs per day. So CE robustness looks good. We lost about 1-1.5% of jobs but this was failure to submit because of the

??: Appreciate why would want to run 2 CEs in parallel. But there are upstream issues with publishing….

There are other clients for submission. Traditionally ATLAS has used Condor for managing submissions.

Use of CREAM in CMS (Ian Fisk)

Issues with an ICE bug emerged – job control was often lost but this is not a critical issue.

Egor works with CMS but is employed by OSG.

GM: So you did not see any overload of the CE?

IF: Doing the gLite WMS submission. Running at a realistic scale. Production jobs run – several hours and run on individual files.

PC: We have just started direct submission. Have not seen any major issue. Most Ces now in production WMS. Some sites there seems to be more pilots going to CREAM CE over LCG-CEs which we do not understand.

WMS sends more jobs (factor of 2 or 3) to CREAM than LCG-CE.

CREAM – Status and plans (Massimo SGARAVATTO)

Old style patch – means release of patch certified by SA3 people. But with new style the developers certify the patches.

Would implementing "sandbox transferring" solve the ATLAS issues?

MS: Also need changes within Condor.

JT: What is the LCG-CE coupled with CREAM-CE problem - maui client server mismatch

DvD: If maui is built to different version... there is a key hidden in the rpm and the key was changed between rpms.

JG: What is the timescale for the LHCb request? For 1.7?

MS: Yes. If we can drop ARGUS to 1.8. Need to discuss further. If it is high priority… 1.7 would be May.

JT: You mentioned the end of EGEEIII. What are the resources for continuing this work?

MS: EMI – was evaluated as 12.5. Very likely to be approved but without major cut. CREAM support is in EMI.

MJ: Remark – Way YAIM publishes… proposal from ST several years ago. APEL is expecting to find an entry for name matching CE. Currently have cron job on mon box to workaround problem.

JG: APEL people need to talk to Stephen Burke and Steve Traylen.

SB: May be fixed in the latest release anyway.

WLCG Tier0-Tier1 Service Coordination Meeting Update (Maria Girone)

JG: Have all the T1s being attending?

MG: Yes all but one.

JG: There are a couple of areas where there is a duplication danger. Eg. GGUS tickets. Also much of the work seems CERN specific. How many people does Lyon send?

HF: One DBA goes and I read the minutes.

GM: Part of it is CERN specific but I find it useful to follow up.

JT: SARA people go.

RT: I dial in …. Not too concerned with CERN specific part.

JG: Not too surprised that the database are well engaged given where it came from.

MG: Dario's comment was general.

KB: If it has the flavour of being CERN orientated then it is wrong. There is another meeting for us to address CERN problems.

JG: What is the status of the TF?

JS: Summary of the items for discussion in the technical forum.

BiG Grid Virtualisation study

LHCb pilot jobs themselves change perhaps 1-2 times per year. It is the job environment that changes much more often.

Grid Cloud and the EES (Mishca Salle & Oscar Koeroo)

JG: The bit about integrating into batch systems – it has been done by a number of teams. Standardisation in this area would be good. Policy and trust issues for images is important. How much can you sandbox what is happening on the machine?

OK: There are a few ways of launching a VM.

JG: Yes, but what about limiting control? How much can the framework protect them?

OK: Most of these systems are not capable in this part. How to bridge the networking how to establish a DMZ. What connectivity do you use. Do you have different policies than those used for WNs today.

??: Can have lightweight… easier to sandbox machine than job.

JG: In terms of ports?

??: Also depends on VM. If you trust the internal OS. If you use briding it is harder to investigate network traffic. Principal – could make certain queues depending on how much you trust the VO.

JG: The term firewalling may be better term than sandboxing.

GS: What we don't trust is the user code. Not trying to sandbox the pilot job framework but rather the user code….

PC: CERN VM does not let users build their own machine.

GS: Users patch their own code into the VM.

Issue is that untrusted user code gets pathed into the VM.

JG: Several HEPiX sessions in Lisbon.

IB: The issue is traceability.

GS: If we got this right, we have one image that is trusted for the environment, and then the user code gets mapped into this an does not require network access.

PC: Many people talking about different things.

TC: The working group has presented things in the past. Element of transmitting images..

Middleware (Markus Schulz)

MS: We never distributed xroot directly.

MJ: We tested 2.1.3 in GRIF.

JG: What does support mean in slide 6?

JG: What updates are in the VO cards?

MS: The roles that need to be supported.

MS: ETICS is now allowing developers to create source artifacts. But the workflow has binaries built from same source but since this is a workflow engine you can have differences.

FH: What stopped working?

If you put VOMS on – something on authorisation then stopped working for that site. VOMS client. It was not even stupidity involved. Just no clear way to understand if you want to package where to pull the sources from.

With respect to other repositoties – like Fedora. Do you have suggestions on how to approach this?

MS: Do you come out with other approaches and be there first – for EMI product stack. Or, look at gLite consortium (gLite contains more than EMI).

JG: History so far not good in supporting many OSes. People will always be building on other OSes.

MUPJ (Maarten Litmaath)

JG: Two distinct cases…

ML: Yes the code could distinguish - glexec could not come to positive or negative conclusion then could still go ahead if normally the payload would have been authorised.

Oscar: On the CALTECH issue, could work with condor group to create an lcmaps plug-in. On Wisconsin problem, that team did not know that they have the same problem as Lyon – to do with AFS and options at submit time. Offer them storage rpms to allow glexec to be moved to different area of filesystem. Same for VDT. So, we could probably tick off this incompatibility issue.

ML: Does condor allow job description to be updated after the fact?

IF: For CMS, we have supported VOMS groups for local countries… so we have allowed them to do this but we have not made any central changes to workload systems. All the work had to be done by the groups wanting this to happen. So, CMS could look at what the other experiments have done. Have other higher-priority tasks.

WMS does little about allocating ..

JG: Does roles and groups information get out through VO views?

JT: Can sort of figure it out based on the estimated response time. Pilot jobs – doesn't matter because this won't be seen. Would then have to do something in central task queue.

CG: For CMS, we said for the current year and perhaps the next log only may be okay.

GS: Not an ATLAS policy to disallow right away and the forum is quite busy.

JG: I'm happy about the transition period being relaxed. Those in log mode can look at the experience of those in setuid mode.

JG: 11 CEs (3 at NIKHEF)... thought about 6 sites ... Fermilab have been running glexec for 2 years.

IF: About half were fully functional.

JG: Just telling sites to install it has not made a lot of progress. But we do need more sites to volunteer. ATLAS tried this at Glasgow. LHCb have tried at 3

MS: swat tests – need to correlate

JG: SCAS has a longer history. Number of staged rollout sites have ARGUS.

ML: There is now a Nagios test (more than swat) glexec actually called. 2 sites that passed.

Oscar: Is the ops VO setup in the US as it is here? Can we test there?

JT: Different testing framework.

MS: But we setup to

MS: We should turn this into a milestone for Tier-1s. Including the publishing of the compatibility.

JT: From backend perspective you are not supposed to tell what is behind it.

IB: Should not specify backend but need a date. So sites need to decide if

RW: Central PAP will be run at CERN for testing and will be available soon for production. This is for the ARGUS clients.

JG: So in 6 weeks would expect to have a larger number of sites showing up with glexec. Can the nagios tests be run as another VO?

MS: Complicated as need to add support for multiple proxies in job.

ML: If you really want to emulate a pilot job you would need to do this but you could also run as yourself.

MS: Every VO that wants to do this needs to change their VO ID card. Tests are not in the old SAM framework.

Target 6 sites and all Tier-1s within 6 weeks.

MS: Who is responsible?

JG: Action is on the Tier-1 sites.

FH: We would not be able to do this.. we need to get familiar with SCAS again

IB: Should be before next Overview board.

IB: Ask the Tier-1s to each tell us their timescale and then turn this into a milestone. Milestone will be that the tests pass and everything is properly configured.

OSG update (Ruth Pordes)

JG: I had a discussion on installed capacity with Lawrence but I will need to check the details and get back to you.

DK: You seemed to be surprised that CERN does not have 24hr cover over the weekend.

No commitment from CERN or EGEE to cover this but people do their best and probably will cover it. But anyway it is best to use generic mail list and definitely report when you can even at a weekend.

JG: Sounds a bit like the situation with tickets where the tickets were held up by the process even though the site admins would often have responded.

AOB:

JG: Ruth has asked if the GDB would like to visit a US Tier-2, perhaps next year? This seems like a good option – let me know your views.

Thanks to the local organisers.

Meeting ended at: 15:25.

[10:38:37] MIhnea Dulea joined [10:38:40] Phone Bridge joined [10:38:41] Tiziana Ferrari joined [10:38:41] Claudio Grandi joined [10:38:42] Gilbert Poulard joined [10:38:42] Jeff Templon joined [10:38:43] Andrew Elwell joined [10:38:43] Tony Cass joined [10:38:43] Daniele Bonacorsi joined [08:16:03] **John Gordon** Casn anyone remote hear from Amsterdam? [08:21:44] **Oscar Koeroo** Hi, audio is OK [08:57:02] **Tiziana Ferrari** I can't hear anymore [08:57:37] **Tiziana Ferrari** can others hear? [08:57:41] **Claudio Grandi** no [08:57:57] **John Gordon** We just reset the polycom in the room so we may have dropped out. Back soon. [08:58:37] **John Gordon** hhh [09:02:56] **John Gordon** Is there anyone else in this chat room? [09:03:31] **Oscar Koeroo** yes [09:03:49] **John Gordon** Can you hear from the meeting? [09:03:55] **Tony Cass** I can hear fine [09:04:03] **Oscar Koeroo** low volume, but I can hear Romain [09:04:10] **Tony Cass** but I had to increase the gain locally [09:21:26] **John Gordon** 30 min coffee break. Start again 1050 - 10 minutes ahead of agenda. [09:54:43] **Stephen Burke** I don't see the slides on the agenda? [10:38:44] H.323 User joined [10:38:46] Oscar Koeroo joined [10:38:48] Marco Bencivenni joined [10:38:48] Richard Gokieli joined [10:38:48] Stephen Burke joined [10:38:50] Mohammad kashif joined [10:38:50] John Gordon joined

[14:17:39] John Gordon joined [14:17:42] Ruth POrdes joined [14:17:42] Phone Bridge joined [08:16:03] **John Gordon** Casn anyone remote hear from Amsterdam? [08:21:44] **Oscar Koeroo** Hi, audio is OK [08:57:02] **Tiziana Ferrari** I can't hear anymore [08:57:37] **Tiziana Ferrari** can others hear? [08:57:41] **Claudio Grandi** no [08:57:57] **John Gordon** We just reset the polycom in the room so we may have dropped out. Back soon. [08:58:37] **John Gordon** hhh [09:02:56] **John Gordon** Is there anyone else in this chat room? [09:03:31] **Oscar Koeroo** yes [09:03:49] **John Gordon** Can you hear from the meeting? [09:03:55] **Tony Cass** I can hear fine [09:04:03] **Oscar Koeroo** low volume, but I can hear Romain [09:04:10] **Tony Cass** but I had to increase the gain locally [09:21:26] **John Gordon** 30 min coffee break. Start again 1050 - 10 minutes ahead of agenda. [14:17:43] H.323 User joined [14:17:44] Christoph Grab joined [14:17:47] Daniele Bonacorsi joined [14:17:48] MIhnea Dulea joined [14:17:48] Jeff Templon joined [14:17:48] Claudio Grandi joined [14:17:48] Mario David joined [14:17:49] Stephen Burke joined [14:19:40] Phone Bridge left [14:29:33] Pete Gronbech left [14:44:32] **Stephen Burke** Maarten, you're very loud! [14:50:19] MIhnea Dulea left [14:59:39] Christoph Grab left [09:54:43] **Stephen Burke** I don't see the slides on the agenda? [10:57:08]

Stephen Burke Does anyone know if it's possible to get indico to display the timetable in the old agenda format? The option to change the display seems to have vanished ... [11:12:40] **John Gordon** I think it was removed earlier this week. [11:16:00] **Stephen Burke** I can still see other agendas with the details of the sessions, and with an option to change the layout, including the agenda style - is there something about the way the GDB agenda was created that stops that? [12:37:29] **Stephen Burke** Did NIKHEF just leave? [12:37:38] **Tony Cass** it's all gone very quiet... [12:37:53] **Stephen Burke** Maybe they're talking about us

[12:38:27] **Tony Cass** nah, we're not worth it [12:39:29] **Stephen Burke** I see the GDB is in a room with a window - luxury! [12:39:31] **Tony Cass** NIKHEF hasn't been eaten by a black hole, then [13:29:01] **Tony Cass** Are the slides on indico??? [13:53:19] **Ruth POrdes** hello - is the meeting over? [13:53:53] **Tony Cass** Nope.at the big grid virtualisation session now [13:54:01] **Tony Cass** I have video and audio [13:55:10] **Ruth POrdes** I have video but very slow. the phone bridge claims to be over? [13:55:40] **Tony Cass** ruth, you are transmitting sound [13:57:38] **Ruth POrdes** not any more.. [13:57:41] **Ruth POrdes** I am not hearing sound [13:58:04] **Ruth POrdes** what is the evo phone id please 1596661 says it is finished [13:59:06] **Andrew Elwell** 1741551 [13:59:06] **Ruth POrdes** reconnected. better now [13:59:43] **Andrew Elwell** (top right hand 'i' button has it) [14:09:41] **Mario David** argus 1.1 server part is on hold [14:10:27] **Mario David** we will wait for the argus client 1.1, since 1.0 and 1.1 are not quite compatible [15:06:08] Claudio Grandi left [15:08:23] Maite Barroso joined