GDB February 11th 2009

Present on VRVS:

Andrew Elwell

Gabriel Stoicea

Alvaro Fernandez

Mihai Ciubancan

Matt Hodges

Mihnea Dulea

Jukka Klem

Brian Davies

Richard Gokieli

Jose Hernandez

Ron Trompert

Martin Bly

Teidir Ivanoaica

Juergen Knobloch

Andrew Smith

Peter Gronbech

Mario David

Stephen Burke

PM:

Alessandra Forti

Ewan Mac Mahon

Introduction (John Gordon)

Feedback is to keep the GDB on a Wednesday. It was asked if the MB could move to be after the GDB on Wednesday.

IB: Conclusion was that we would coordinate the agendas to ensure topics addressed in the correct order

JG: Not clear on the reasons for not being able to move MB

Book for March meeting!

Currently all meetings at CERN. On June 10th several things are happening so may be an issue with chairing.

LHC schedule:

IB: Likely scenario first injections September with collisions end October 09. Physics run from November for almost a year. Similar to plan for 09-10 run. Short stop over Christmas. Heavy ions. Need to go back to 09-10 resource plans (09 resource for September) and ensure followed and 10 resources in place by April next year as previously agreed,

JG: Sites will be running when they have to install new capacity. What you could not test in advance is scaling – power or network.

IB: Experiment models changing during shutdown (3 months for reconstruction)

JG: If those changes mean need more resources to do reconstruction say then unlikely to get them

EGI: JG: Main thing that seems to be missing is detail around what is an NGI.

JS: Does the new schedule have an impact? EGI supposed to take over is in the spring next year – that would mean a major discontinuity next year. So, what we need has to be in place for this July.

IB: Statement has always been that there should not be any discontinuity.

KB: Why would this be disruptive?

JS: EGEEIII stops and EGI does not yet exist. Just saying that we need to be aware of it.

IB: If there are services run by a ROC that WLCG relies upon then need to make sure they are covered. Next overview board is on 23rd February and can be discussed there.

JS: There are people at CERN not funded after EGEEIII.

JG: That should be well known – know what needs to be done but perhaps do not have the people identified to do it.

File Caching on WN Disk (Graeme Stewart) [ATLAS pCache]

Issue with conditions data held in SQLite file – file became very hot during reprocessing and this led to failures over Christmas. One solution is to cache files onto the WNs – proposed mechanism uses pCache.

Bottleneck in future with 8-cores… do you have a prediction of load on this node.

GS: Should be no worse as currently files are all downloaded

JG: With and without cache what is the load on the WN?

GS: May move to a situation where there are i/o problems but this is the way you move forward.

Wrapper around the file stage command that just checks if file is already cached.

JG: What about caching only some files?

GS: Need to ensure files

JG: pCache runs when?

GS: Once per job but uses a locking system so file only downloaded once to a node.

JG: Implementation?

GS: Currently implementing in production system – will be an option that we can switch it on/off. The config information we need the site has to tell us.

LdA: Not useful for virtual machines. Is there a way to access a physical machine copy? We do not support all LHC VOs and have additional VOs so want to be able to run on other OS and environments.

FH: All T1s have deployed expensive Oracle clusters and this problem should be solved by the Oracle implementations. Why go back to this route?

Dario: Scheduled reprocessing will not hit Oracle.

FH: You are assuming that the Oracle we have can not handle 2000 jobs starting at the time?

DB: The Oracle route is only available to T1s, T2s may also wish to do reprocessing. Transaction level can be hundreds of thousands.

JG: Is anyone looking at Oracle optimization and solutions?

DB: IN2P3 has solved problem using replication in dCache but not all sites do this.

FH: Why are the Oracle databases not coping with this load. Should answer this before implementing another solution.

JG: Talked about T2s. Also, not all the shared data is in Oracle. Anyway, I thought Sasha was looking into database issues.

GS: His group is looking but this way will help too. Nice thing about this is that it is very simple and a big win for efficiency. Just need to say something like can use 10GB in /scratch – does not need a lot of configuration. There is a small deployment cost.

??: You know that sites can wipe the cache at any time?

GS: This cahce is controlled.

??: Are the other experiments looking at similar solutions?

GS: If ATLAS has a share of site resources then … minimum useful cache may be 5GB. This is to help sites… if sites can manage with out it then fine (but their SE must be able to cope with the multiple hits).

DB: Current tar file was of size 2GB. If split by run period then could be files of 1GB per run.

JC: Is there an implementation of this already?

GS: It has been used at one T2. Would test more before wider deployment.

Storage (Graeme Stewart)

JG: A lot of curation data – getting it stored reliably on tape and meeting rates from CERN. All sites shown this work but now find a lot of copying/staging for copying to other T1s or T2s. For reconstruction bringing data online is a big hit. Asked experiments questions on this…

Tests of bulk pre-staging. A repeat of tests done in Autumn of last year. Want to be able to reprocess a year's data in a month.

GM: In the previous tests ATLAS had 100% of resource share. In a retest it would be better to say cap use at 50% (for PIC) which is the ATLAS share.

JS: These tests were

Xavier Espinal: Looks likely that we can do better than results…

Overloaded NDGF. Rate was 40MB/s.

JG: Do they have one dCache headnode for all their sites?

GS: They perhaps did not have the power required. Library not performing as expected. It is thought to be fixed and we need to retest.

JG: May be that other sites have been through this for CMS….

We asked for 9000 files and got 3000. I believe the backend tape system should be able to handle rates but not seeing it because of issues in the upper levels.

For CNAF, some glitch has crept into the system since last year.

LdA: Same version as last year. CASTOR support team tell us known bug and we should upgrade. Finding 64-bit hardware. All the same, difficult to understand how the same setup worked last time.

BNL still to do.

Lyon – problem HPSS and dCache interface. Currently all requests are passed serially so performance was predicted to be poor.

FH: Brookhaven is using mediator and we are looking at it. Takes and reorders requests. We have to modify it for our needs.

What is MIA?

GS: Missing in Action. Some files disappear – get back from buffer but not on to disk.

IB: Were any sites writing at the same time?

GS: No, at least not writing ATLAS data. We would like to move forward to the realistic data scenarios.

TC: CASTOR fastest and slowest. It is the number of drives behind the SRM. How many drives behind these figures …

GS: RAL had 10 drives and all seemed to be in use.

JG: Have seen tape drives working at line speed for CMS.

GS: Also need to worry about packing – by physics tasks or is that too much granularty?

FH: Can we trigger the bring-on-line early?

GS: Model is – reprocessing task defined once everything locked. The production system then asks for data… as data comes back the production system launches jobs for that data. Waiting for 5 minutes for first jobs to run… do not need all data online to start.

FH: Are there alternative solutions? Pre-staging. Working on longer term but if we can find something for the interim we wil

JG: Bring on line and then copy to the WN? So few 100MB/s is easy?

GS: Yes if have enough disk servers – it is an infrastructure issue. Know that some RAID arrays read and write very fast but not if doing both at the same time. Do not have internal knowledge to fill servers in certain way and anyway then you would be running jobs off single servers.

KB: The table slide… we conclude that for the March reprocessing challenge will use at least RAL, PIC, Triumf and ASGC… and the other sites can they be ready?

GS: Given the volumes, SARA is probably fast enough. Deadline for decisions is start of March since need 1 wek to clean

LdA: Preliminary analysis is that problem is in the SRM layer. So strongly advised to upgrade and will do this in the coming days. Can the tests be repeated?

GS: Yes, Claudio ran the tests.

KB: FZK not present, but the are doing an upgrade now – some pnfs thing.

JS: Second bullet for overlap, doesn't this have to be a milestone?

GS: Yes

JG: The tests can be run on demand which is useful. Thought CMS would only schedule things if in line with their plans. Find out when CMS will be active and then fire the ATLAS on demand tests.

FH: If CMS can provide sites a way of (eg. List of files representative of activities) running their part then sites can do it.

IB: Is it not better to have facility available to sites to run on demand stress tests – they would want to test after changes.

MD: If you want to be in touch with sites as you … GGUS template to push r esults?

GS: We already have the contacts so probably not required.

RT: Late last year got list of SURLs from Simone and this enabled us to run some very useful tests so really advocate this approach to allow self-testing.

JG on feedback from the sites – see slides.

KB: Question about BNL (could not hear question)

JG: Conclusions – sites acheievin rates for raw and ESD but no confidence they all can do it simultaneously.Sharing of work between analysis and production is thought to still be an issue.

This is the basic feedback for the reviewers next week.

HR: ATLAS have demonstrated the importance of collocation.

JG: The sharing of data between drives.

Status of the ALICE :WMS Usage (Patricia Mendez Lorenzo)

The myproxy server can use infrequent delgation but move to glexec will lead to a factor of 10 increase

JG: What happens after the 2hrs (slide 8)

PM: IF have 10 minutes can also have one resubmission which reduces the load.

JG: What happens to those jobs – marked as failures?

PM: At the moment job gets resubmitted up to 3 times.

??: Are you talking about resubmissions or matching?

??: Job is still in queue so it is the matchmaking….job is still waiting.

MJ: GRIF experience. If just using job status, can happen that have waiting without any job status reason, probably the most frustrating for user since waiting with no additional reason.

We have some big users of the WMS and the WMS is hardly recovering.

You can configure the CPU and disk load. At GRIF 8-core machine, not much loaded …

??: There is an internal bottleneck.

??: WMS is trying to mask users from problems on the grid. Would try to push the developers on implementing this behaviour.

Could not follow the discussions.

Classad library problem – issue with the way that the matchmaking is being done.

Do not know why backlogs observed.

JG: I thought a CREAM CE at CERN was on the plan

PM: Priority was SL5. When it has been completely tested they will think about it.

TC: PM correct. Have CE for this once Sl4.SL5 problems resolved.

PM: CMS are having a slow down in job submission but for different reason – input sandbox. ATLAS not using WMSes at CERn and for LHCb do not know.

Other sites have just one WMS so the jdl is very simple in these cases.

JG: So at sites with WMSes themselves you don't see this problem?

PM: The CNAF WMSes cover all sites in Italy.

JG: So the number of jobs going through CERn is a factor of 10 higher?

CMS: True that see instabilities from time-to-time. Also uncover systematic effects like input sandbox. Are you using the WMS limiter? Puts WMS in draining as soon as certain job load reached.

PM: Have asked for this at CERN.

MJ: Limiter does not work – low load on machine but WMS unresponsive. At GRIF there are 8 WMSes. Find that things work ok for couple of weeks and then problems. We now have problem with matchmaking not working at all.

??: Means requirements in the job are not right.

MJ: No, in the new WMS it is exactly the same jobs with the same BDII.

JG: So, the problems seen by ALICE now seem more wide. Is this something to be tackled in SA3 or a team can investigate in SA1?

OK: There is an update to the WMS in pre-production. Otherwise standard process of registering problems and developers taking care of them.

Maite: SA1 can investigate but it sounds like ALICE has done quite a lot of investigation already.

ALICE: We need some timeline because it is a problem. Spotted in January and now February.

??: New WMS is in PPS.

ALICE: Chances that it solves the problem is very slim.

JG: Given that this is a standalone thing why can't you just point at the PPS version?

PM: Well we would need to setup a VOBox in PPS.

Conclusion was to try the PPS version. Work with SA1 to investigate further.

JG: Flavia (unfortunately ill today) was going to report on installed capacity – change info providers to produce reports to compare figures with pledges. Have agreement at MB on the requirements and Flavia was going to give details of the delivery plan.

??: That was agreed by the MB?

JG: Document on requirements was agreed but not the delivery plan document. How you use the glue schema etc., I don't believe Flavia had feedback on those particular areas or that there was conern.

CMS: Trying to do some monitoring for CMS so would find the details useful.

JG: This is to be implemented for reporting and won't be good enough for scheduling or real time operations.

??: Just need agreement on the names of the fields.

JG: Part of the plan was good documentation on deployment, how we support the two benchmarks etc. May send the documentation around.

JCasey: Rob Quick for GOC and JC for EGEE, will provide more details for gstat, gridview etc. and when these could be included in the availability reports. Will work on plans (with timelines) for this in the coming weeks.

LUNCH.

EGEE authZ framework (Christoph Witzig)

Main point of talk is to introduce a deployment proposal. Request for volunteer sites to try the service.

FH: On your sldie 13, what is the relationship between ops auth service and SCAS.Will this replace SCAS?

??: SCAS is the shortterm solution to allow glexec on WNs. Longer term is this solution that will also be integrated into other services, so this is in the longerterm solution.

FH: Are the storage systems going to use this?

..: At the minimum it would be good if they take the list of ban users. It would allow an admin to ban users on CE and SE in one easy way. How the list gets to the storage element that is something we need to look at each case. They may get the list directly or they can call out to the service. This service is typically on the command not on the individual files. You would not go out and query status for every ls command for example.

FH: Can you comment on the alt services – it may be a single point of failure (not read the paper)

..: Appendix describes two ways to make fail safe. Install config on several hosts and the client calls out to each. Or you can replicate the deamon for the decision and admin.

FH: How many requests per second can this handle?

..: we have not yet got that information. From SCAS we are talking dozens of requests per second and this service will definitely handle this. Can not say if limit is 45 or 145 or more.

??: Right now we are implementing the SCAS service and you can see the implementation plan for that. You can test this service in the same way.

..: Should point out that this system does not depend on a shared file system.

FH: Sites should be informed if there is a global banning.

JG: Mentions OSCT banning list but

FH: Users approach site and the site may not know that the user was banned.

..: Should be covered by the security officer already

GS: Can we have a meaningful error message when the user is banned!

JG: You mentioned a few services. Is this the full set of services required for a site to have a single policy on banning. Do we need it in LFC and FTS for example.

IB: LCG position – this should not get in the way of us deploying SCAS and pilot jobs in the way we already planned. This looks interesting in the longer term. People should not stop deploying SCAS and wait for this solution.

JG: Nothing here looks like a big-bang approach. Anyway, is this on the WMS workplan too?

CW: Yes.

Grid Configuration Monitoring on Worker Nodes (Thomas Low)

JG:What is an acceptable time from sites for these wrappers.

Do you want to start counting the WNs… will sites be happy about the structural probing? What is intrusive?

JCasey: There is a view that the stats are useful. Julia – CMS have to size there jobs depending on resource available so the feedback is useful.

MJ: I think you will monitor too many details. CMS may decide to submit lots of jobs just as the site decides to remove some resources… if you want to go into all the details you will

ST: Never validated subcluster information – between what's published and the reality is a useful check.

FH: Sites are not trying to hide things

IB: The intention was not to have more monitoring but to allow help in diagnosing issues.

JC: Tests can run on a longer timescale – for example every 1 month.

JShade: SA1 operations can then see things that help with configuration management.

JG: What are you going to do with the information? Are you raising tickets …

JC: Main reason for job wrapper tests was to understand things when something has gone wrong. Here is something that tells us something about the environment over the last 24hrs. The amount pushing alarms would be very small. It is data mining not alarms.

JG: Concerned about associations…

IB: Main push for this was to help answer the questions about why jobs fail and I think this is a tool that helps you debug that.

NT: It is for use when following up

MJ: The VO should not be using this for scheduling

JC: They may use it to help tailor

MJ: They should rely on what the sit publishes.

..? From operational view. … aha, so it is injected with every job.

JC: It has always been there but not reactivated.

AF: That is because the original monitoring brought down the WNs.

JG: Certainly need to be sure there is no potential for things to hang.

TC: You say running deterministically, will run this to check every node at CERN? That would take a lot of job wrappers and I have a concern

IB: You know how difficult it is to push updates out to sites.

TC: Penalise sites that are reacting because they already monitor closely and this is an overhead.

JC: Average is under 10s for each job.

MJ: Agree with Tony. On most sites the version is consistent.

JC: You cap it at 30s or less if you want to. Is it a problem to run with every job? Or you can request it runs 1 in 100 jobs.

JG: What were you expecting from this meeting? A decision that it is ok to do more or just to inform us of the direction?

Thomas: Well it is on the PPS.

Middleware Update (Andreas Unterkircher)

Rollback.

SA1 request for more sites to take releases as early adopters for specific components.

??: What would you do with the multiple WN versions?

AU: To allow older certified version and newer release on the same machine. Also multiple gfal versions etc.

IB: What is the error rate compared to SCAS?

There is a race condition here! The solution was to kill the process after 30s and restart.

The NIKHEF are doing the work on SCAS and if this solution is not acceptable then they will continue to work on it. The method above is a workaround but not a fix.

MJ: A memory leak in the software can not be deployed like this.

??: An error rate of 0.03664% is acceptable?

JG: It was thought that LCAS was not scalable so shared file system errors are likely to be used. A small error rate like this is not completely unacceptable but having a memory leak issue every 30s is a problem. You may also get a higher rate when a whole farm goes down and comes back up with many jobs starting at the same time.

SNewhouse: Are there independent requirements from the glexec end?

FH: The jobs restarting hundreds of jobs then requests per second of order 10 is a toy model.

OK: 2 calls per job and allowing for peakyness. In the testing at NIKHEF at has been shown to do 20.

CG: SCAS was to be used only on the WN and the restarts issue would go via a LCAS on the CE.

MJ: We should be aware of how the service degrades at higher rates – 30, 50 and so on.

AU: There is a pilot and the current error rate is not unacceptable

JG: Agreed – if nobody has said that this memory leak can not be fixed.

Status of the LCG-CE (Andyrey Kiryanov)

DB: Do we know how many sites have the latest CE?

AK: Almost all of them.

JG: Is there not a plan to deprecate the 3.0 version?

DB: So we have managed with what we have and does this mean we stay ..

IB: Plan is not to switch to CREAM but do a parallel deployment until such time as CREAM is shown to be better.

DB: If glexec integrated with LCG-CE then do we need to move on?

JG: well, we have not yet seen chaotic user analysis.

AK: Almost impossible to do a load-balanced CE with the LCG version. Whereas with CREAM the information is stored in a database and load –balancing will become an option.

JG: LCG-CE with SL5 and ..

OK: There is no plan to port.

JG: Do we know that it would not work?

MJ: Still have to support a lot of users on CE. There is no point to say that we don't need it.

IB: Your point is that we should not put more effort into? It is not time critical apart from SL4/SL5 issue.

DB: Looks like CREAM option is on the timescale of the LHC upgrade!

IB: The other thing is that the LCG-CE can not pass other parameters to the batch system.

JG: There seem to be two competing views and they are balanced. If we can wait till 2010 then

IB: Would like to see some significant milestones for CREAM this year – even before July by when it will be 3 years since we set the conditions for it and it has not got very far.

Pilot job frameworks review – update (Maarten Litmaath)

Q: What is the timescale for the conclusion.

ML: I am not the boss of these people and can only report on how it is going.

JG: What of the patches?

ML: Last year ALICE/ATLAS identity switching was made to work. Getting close to a glexec that is using the SCAS facility. From the VO perspective it should not matter – call the same way. This is more the testing of the glexec/SCAS infrastructure – how does it work at the site. …

JG: I am interested if it integrates smoothly with experiment frameworks.

PC: Glexec returns status of the payload but does not distinguish if user not accepted or glexec not responding.

ML: The developers accept this concern and estimate 2 weeks to fix it.

Some issues resolved – issue of calling with an expired proxy explained. If site properly setup then should always get sensible mappings. Also looking for memory leaks etc.

IB: Coming back to the slide, goal was to convince sites that the frameworks are safe for the sites to deploy.

ML: LHCb ok. For the others there are concerns – there is a lot of code in every framework. We could dive into the code and check every line but have so far approached it in terms of principles of approach (use of secure connections etc.)

IB: This is a confidence building exercise. Based on the recommendations you made so far would you recommend the frameworks?

ML: LHCb is completely ready. But did not look inside the code.

IB: Should the sites have confidence?

ML: CMS model looks pretty good – uses standard Condor. But, only used for production in the US. They originally expected points at FNAL and CENR, but if having at every Tier-2 then more review may be needed. The fewer the dependence on external resources the better – box access limiting…

IB: Can we make the statement that sites should accept jobs then if they come from FNAL or CERN?

ML: Claudio may be able to say more for CMS

CG: Certificate – accept at the submission point. …

ML: For ATLAS – the submission mechanism looks fine if it is implemented as described in the twiki. We knew there were a few desirable features but some of the features are not fully secure – only uncovered yesterday evening and it is being looked into. For ALICE I have been working with them to come up with a first version of the document. There are desired improvements to tighten the security but I would not have a problem implementing what is described now. In any case there is an opportunity to put more hurdles for hackers.

The issue with Panda is independent of glexec.

Will the group continue working after the implementation starts?

ML: The mandate is flexible.

VDT & OSG (Alain Roy)

OSG software coordinator

OK: Comment – working with ETICS for proper source releases. You are an integration provider so you probably rebuild so will it be useful to you? At the moment we have certification and pre-production. If you wanted to take something at the source level, and at the moment you have to reverse engineer the binaries. This ETICS move would make it easier for you.

AR: As much as we can be plugged into this the better.

OK: Are you happy with your build system in this respect.

AR: We use the Metronome build and test system. Happy with build mechanisms… sometimes when we rebuild your software on other platforms we hit the problems first. From that aspect it is challenging.

OK: Would like to ensure the feedback fully benefits the project.

MD: Have you forgotten VOMRS? We found it was too heavy to provide with gLite and it was to be in VDT.

AR: At the moment you could say the rpms come from VDT.

Meeting closed: 16:40.