GDM notes 6/9/04

Attendance: Vicky, Bob, Rob K, Matt, Don P., Julie T., Mark Kaletka, Mike Stolz, Eileen, Mark Fischler, Steve Wolbers, Gerry B, Jon Bakken
Video: Frank Wurthwein, Rick St. Denis, Ruth, Lothar

Introduction from Vicky
> Please have patience in assisting this series of meetings to become effective, don't give up immediately. Participate – don't read email.   We will work to get slides ahead of time and make logistics smoother, but please give it a chance even in advance of that.

Expectations from CDF: Frank Wurthwein

> The primary reference for CDF expectations is Section 9.4.1 of CDF plan for Bird Review (http://hepweb.ucsd.edu/fkw/cdf6640_computing_plan.ps)

> Distillation of requirements from this section:

>> Require basic SAMGrid/storage functionality plus: Quota-enabled permanent disk storage mechanism: flag to SAMGrid/Storage services/Privilege project to address that requirement in next week's meeting.
>> Concatenation requirement: need tools for this
>> Quota-enabled tape storage of user-generated analysis files.
>> VO policy implementations across remote centers – where remote centers may use local storage or remote storage in conjunction with local cpu

Expectations from CMS: Lothar Bauerdick

> Good timing for the forum to start: in recent weeks, have a clearer picture related to OSG (recent meeting at Chicago); have a process to assess requirements; on LCG side, have a new communication structure in place

> Have defined CMS/USCMS milestones to move distributed computing from Grid3 to OSG, and to have more specific functionality for analysis. These are related to demos of technology still under development – will include 'service challenges'. The service challenges will provide opportunities to work on the system aspects using a long term view of building their functionality, instead of a short term view of meeting a physics goal by whatever means possible. For example, the Tier 1 centers will test 500Mb/sec throughput to CERN sustained – not coupled to physics milestones.

> The projects already started are Runjob and Privilege & Authorization. For Runjob, there is not a complete work plan yet, but there is a definite set of expectations from CMS. For Privilege & Authorization, the functionality is clearly needed by CMS and there is a milestone at 1 year.

The following concrete CMS milestones will feed requirements into the Grid & Data Management projects:

10 million events/ month fully simulated & reconstructed & provided for analysis (compare D0 1 million/month steady state, CDF 92 million events in one month.) - this milestone is due now

Be capable of running analysis job on Grid: Nov 04 (I did not get specific scope of this milestone).

Want to see joint projects on storage services in context of OSG, common storage elements based around SRM. Permanent storage, transient storage, making a lightweight storage element available in a deployable manner.

Blueprint – architecture effort beginning in PPDG – need to understand where CD effort will fit in this.

Interoperability effort important – grids & storage services.

Q. Frank – What about Replica Location Service? A. That is missing functionality. Bad performance so far.   Where is RLS in the architecture – under discussion.

Expectations from D0: Amber Boehnlein

Runjob –
            would like clarification of project structure & management; please address in   future meeting
            Drafting expectations at Fresno (this week)
            Request to visit
Integration of various tools – specify interfaces
SAMGrid
- Security / VO man
- JIM for MC
- Jjim on Farm
- JIM for RECO by Spt
- 2nd gen monitoring
- LCG interoperability
- Tape migration
- SRM interface
- JIM on FNAL analysis platform Mar 05
- Program to reduce operational load
- Full Linux operation Mar 05

Storage services
- SRM interface for enstore,dcache,HPSS,sam (disk) cache
- Tape migration (next generation)
- Dcache for staging data to remote sites Jan 05
- Dcache for online end of shutdown (online machine upgrade, operational issues)
- 2nd gen monitoring enstore & dcache

Additional
- calibration db proxy servers
- luminosity db server

Facilities
- test platform in GP farm to be augmented or separated??
- FNAL grid resources?

Q. Frank Is FNAL grid platform separate hardware or portal to existing resources e.g. CAF? A. That's also what Amber was asking.
Q. Frank Why do you need SRM to enstore? A. Not all tape access at D0 uses dcache.

Frank: CDF had not mentioned Runjob or LCG – will return to that here. This is because cannot find CDF effort to contribute to Runjob. Asked for this in PPDG supplement proposal. LCG – are pushed by offsite collaborators to not expand except into LCG resources.

Vicky: LCG is not a show stopper big deal, right? A. Yes, there is a lot of progress on this, so it should not be a huge burden but it IS an expectation.

Division expectations I: Vicky:
My expectation is that FNAL will move – eventually – to common resources for all experiments. Won't imply losing control of spending money, won't imply complete uniformity. Implies level of interfaces and commonality that is unprecedented. Will be part of instructions to projects to consider and accommodate this. Will be presenting a large set of resources on the Grid – will serve Division and science better.
   Clear that we don't have enough people to do all we want to do. Could benefit from twice as many people. Can't waste resources or be frivolous. Want to put resources where they will be well-aligned. Buy into doing good projects and getting them accepted. There will be arguments – will be hard work to prioritize. Will need to meet milestones driven by experiment and some by the political considerations. Together will make fantastic progress if we can do this. Grasp opportunities to work internationally.
SRM, SAMGrid, CMS work, and GRID3 very successful – get our money's worth.

Division expectations II: Ruth
Need to use these meetings for generating actions and follow-ups. There will be work between the meetings as well, for CD projects and for stakeholders.

Comments from other Division stakeholders:

CCF Don: Facility aspects to these projects: e.g. facility stance on security, storage, networking.   These should be addressed in this series as well.
Comment Amber: Put this in the third meeting, make discussion a fourth meeting.

CCS Mark Kaletka: Facilities in a broader sense – helpdesk, farms, databases in distributed environment. Include these as well.

CEPA Mark Fischler: Most important thing heard was Vicky's statement – watch out for blind alleys.   How to coordinate with multiple experiments, but not falling into that trap?
Comment Wyatt: This forum is intended to be some help with that, in the sense of providing a way to make mid-course corrections in projects based on changing expectations. It is not to be expected that we never waste an FTE hour!

EAP Steve Kent: Don't want to get snowed under.   Right now Grid is not critical path, so can make a contribution because can take some risks, be guinea pigs.

Action items/requests.

1) Reschedule 3rd meeting to address facility requirements.
2) D0's clarify runjob management – how/when to follow up
3) CDF's address quota-enabled disk storage
4) D0 address grid testbed and grid resource deployment
5) Discuss evolution of metadata services

Q. Frank : Is computing infrastructure for Run II and LHC identical after, say, 2006?
The only problem he sees is metadata – file catalog not being separate.
Discussion with Rick St. D: there is an ongoing metadata group in GridPP addressing these questions.