GridPP Technical Assessment Group of the TB

Overall Assessment of Experiment Applications

Status of this document: FINAL 9-Nov-01

This document represents the final version of the technical assessment document, and is declared final in order to have it in place by 8th November (the soonest possible after the collaboration meeting)

1. Membership of GridPP Technical Assessment Group

P.Clarke (Chair)
D.Colling (Job submission and SAM)
A.Mcnab (Testbed components and BaBar mapping to EDG)
R.Middleton(Information services and DataGRID overview)
T.Doyle/G.Mccance (Data management)
J.Gordon (Mass storage and local infrastructure)
L.Robertson (liaison to CERN fellowship programme)

2. Structure of document

In evolving the work of the TAG is has become very clear that the assessment of the experimental application proposals needs to be split into three categories. Accordingly the document is structured into three sections:

1. BaBar,UKQCD,UKDMC,MINOS (i.e. the non-LHC and non-FNAL experiments)

   With the exception of BaBar these experiments are relatively small, but importantly all are relatively independent of a large more pervasive environment. As such the documents submitted to the TAG on the 8th were in the correct form and it made sense for the TAG to assess the applications in terms of individual Grid components, to provide feedback, to identify possible commonalities, and finally to make recommendations on resources needed to

realise the applications.

2. The FNAL experiments

   Following an early assessment of the D0 documents, and detailed discussion with one of the architects of SAM, it became clear that D0 is already tied to a working Grid-like framework . SAM has a fairly clear Grid upgrade path, and thus it made sense to assess this in a more holistic way, and it certainly did not make sense to assess it in terms of individual EDG components. SAM is part of the ongoing PPDG programme.

   CDF are still in the process of assessing how best to build a distributed Grid application which fits with FNAL CDF policy. The most likely scenario would be to adopt and adapt SAM and thus CDF is assessed under this assumption in order to identify commonalities with D0 and the prevailing FNAL Grid infrastructure.

3. The LHC experiments

   The LHC experiments are fundamentally different. It is accepted that Grid posts assigned to these experiments must rightly fit into the wider and longer term experiments planning. To put this another way it would be absurd for GridPP to in any way be driving a very different agenda. Thus all such posts will per definition be contributions to a full programme.

   On the other hand GridPP does need to see clear and identifiable milestones in the development of true experimental Grid application - i.e. built upon existing Grid components in the next 1 and 2 years. Thus LHC experiments are assessed in terms of a suitable contribution to realise a programme against tangible milestones which can be monitored.

The document also includes a brief assessment of support for areas of middleware which should be found in an ideal world to fully underpin UK responsibilities in DataGRID.

The document also includes an initial assessment in detail of the areas which need to be supported and augmented at the Tier1 / Tier-A prototype centre located at RAL. Some thought is also given to the needs of Tier2 centres (but this is preliminary and intended as a marker that ideally effort would be found to directly resource Tier-2 centre operations for GridPP)

1. Non LHC/FNAL experiments

UKQCD

The UKQCD Grid requirements described are clear, limited and simple in phase-1. They seek to set in place a UK wide data storage and retrieval system, using meta data to catalogue and access the physical data sets. All data is moved to the local site where a non-grid like job submission occurs. This is a well founded Grid application which should be completely viable using EDG-TB1. Phase 2 is less well defined today, but will certainly add Grid job submission to remote sites. Again this should be completely viable using EDG components.

The updated submission on 30 Nov adds timescales and explicit manpower requests. The milestones given are optimistic given the in practice start date of the FTEs to be allocated, but are otherwise fine. They are more than sufficient for TB requirements of monitoring.

Resources: Taken in isolation a minimum of 0.5 staff post would be required to configure and support the experiment specific aspects of Grid data management for all sites. A further post would be required to build substantial applications on top of this.

This effort would have to be placed at sites with existing EDG deployment experience (e.g. Groups with associated LHC experimental Groups.) to benefit from sharing of infrastructure and expertise.

We also note that there is great overlap with the UKDMC/MINOS/ANTARES application, and therefore collaboration is strongly encouraged.

UKDM/MINOS/ANTARESC- Data Access

UKDMC/MINOS and ANTARES make it clear that they do not have huge resources to devote to Grid at present, but would like to benefit from Grid tools which exist. Accordingly they propose a simple and modest Grid framework, which will focus upon integration of metadata, replica management and data transport. This follows very sensible discussions which followed from the original document submissions.

They have presented a straightforward proposal for this which should be easily implemented based upon EDG TB1 components.

Resources:

Taken in isolation a minimum of one post is needed to configure EDG-TB1 components for all experiments, ideally 1.5 to allow more fully for the experiment specific parts of the work.

This effort would have to be placed at sites with existing EDG deployment experience (e.g. Groups with associated LHC experimental Groups.) to benefit from sharing of infrastructure and expertise. These experiments have already noted this in their submission.

We also note that there is great overlap with the UKQCD application, and therefore collaboration is strongly encouraged.

UKDMC/MINOS – Remote Control

A truly different Grid application from most others which arise in HEP experiments. Within the context of EDG no thought has been given to this. It clearly has elements of security and real time criticality.

This doesn't obviously build on EDG components or architecture (although with some more thought some synergies may be found). As such it would need a lot more work to define a coherent programme as the real time nature, critical processes, security etc lead to a context which is rather different a priori from a data oriented Grid

To develop the whole thing alone would clearly take many FTEs. The TB cannot recommend that GridPP undertake such a project in its entirety, as it is outside the goals of a Data Grid.

However as was noted some other Grid projects (Collaboratory) and the CLRC eScience centre have interest here. There may well be overlap with AstroGrid remote control applications. To take this forward MINOS/UKDMC would have to explore possible collaboration with these other projects/bodies.

BaBar

Details of the BaBar experiment assessment are still not done. This is an inevitable consequence of the massively aggressive timescales for getting documents in place, and is not a reflection on BaBar. This will be done in conjunction with BaBar prior to the review meeting and presented as an addendum.

2. FNAL Experiments

D0

The D0 infrastructure is based upon SAM. SAM is a working distributed framework which already incorporates some Grid like elements. SAM is neither a simple atomic Grid like component (e.g. a storage manager) nor is it a complete Grid application.

Instead it is somewhere in between offering a framework (including a command line interface) to which the user can identify data sets via a metadata specification interface to a relational database, submit a job, and which then takes care of remote physical data access and subsequent job submission upon the locally staged data.

SAM will be progressively upgraded to incorporate more Grid components. The document lists these steps in detail and we consider them to be very sensible and incremental, and to lead to a much more Grid aware version of SAM which matches the goals of GridPP.

D0 has specified a Grid user interface application to be built on top of SAM. This will provide a Monte-Carlo production system and user analysis interface.

The TAG questioned whether this work on SAM would be done in collaboration with PPDG and Globus in a way which would ensure that SAM evolves to use emerging world standards (as opposed to using experiment specific / non re-useable solutions). We were assured by the GridPP EB representative that since D0/SAM are integral parts of the PPDG programme, and that the SAM architects were working with Globus, that this would be the case.

Finally we note that the problems of interfacing Kerberos to Grid security are ongoing, and will have to be resolved as a prerequisite to much of this programme of work. We understand that this is in hand at FNAL.

The document submitted on the 30th.provides some good deliverables (which might have slightly over ambitions timescales , but this can be refined). As presented the deliverables are a little lacking in specifics, but if they can be taken in conjunction with the specification of SAM enhancement given earlier in the document are more than adequate for monitoring through the TB.

Resources:

- A suitable contribution to the underlying work on further Gridifying SAM is 2 posts. This should be seen as core work (not D0 application specific) which might benefit CDF as well (see below). This assumes that these 2 posts will add to other resources provided from other sources.
- In addition D0 would require an additional post develop the Monte Carlo production system to be built on top of SAM.

CDF

CDF have presented a very clear and detailed document outlining their plans, and a suggested set of timescales and goals for building a Grid environment to support CDF physics in the UK.

CDF present a programme which starts with a manual implementation of non-grid infrastructure (a pre-requisite to GridPP support) and then foresees development of

1) A Grid framework for data management, coupled with by hand job submission, with a target of 1 year

2) An augmented framework using Grid resource discovery and job submission on a timescale of 18 months to 2 years.

The lists of requirements and use cases as well as timescales is extensive and detailed, allowing the TAG to gain a clear view of the goals of the CDF submission.

Formally the stated goals, at least within the UK, would seem to be achievable using EDG infrastructure and components. However we believe that this is unlikely to be a useful way forward in the long run as it would probably de-couple CDF from any central infrastructure supported by FNAL (note: we distinguish clearly between Globus alone, which FNAL and CDF will use, and the value added infrastructure which is provided by the EDG testbed releases). Whilst this remains an option, we feel that the most likely way forward would be to adopt the SAM infrastructure, and thereby become part of a wider FNAL structure.

CDF-UK are currently evaluating exactly this question in order to find the optimum way forward.

For the purposes of this document we assume that they will adopt SAM, and we comment accordingly. This would appear to gain the advantage of being able to make use of GridPP effort to benefit both D0 and CDF in a more efficient way.

In order for CDF to adopt and make use of SAM requires CDF specific effort in several distinct areas:

- Import, deployment and support of SAM stations in the UK
- Adaptation of parts of SAM to function with CDF. This will almost certainly include adapting the metadata catalogue and replica management to match to the CDF specific data model. This means assessing how to interface the existing data base to SAM services, and consideration of a possible distributed database. There may also be other SAM components requiring some work.
- Implementation of a Grid user interface application on top of SAM.

This is all assumed to be on top of the core SAM Grid evolution work as detailed under the D0 heading, and an assumption of solution of the Kerberos problem.

The addendum document submitted on the 30th Nov provides the necessary milestones which are adequate for the TB

Resources:

- CDF request two posts to work on adapting SAM for the CDF environment. This would need to work closely with the core SAM team. It is difficult to estimate this any more accurately, and the TAG certainly agree that it will require at least 1 post.
- In addition CDF would require an additional post develop the GUIA as well as configuring and managing CDF specific information services including resource advertisement and management policies. This assumes some effort

already exists in CDF to work in this area and that this post would augment this.

3. LHC Experiments

ALICE

Alice Grid development is being spearheaded by groups outside the UK. In particular those groups developing AliEN.

Given limited resources the ALICE UK interest has not been able to devote any resources to explicit involvement in the Grid, and does not at this stage propose to undertake a significant role from the UK in this venture.

It would not make any sense for GridPP to try to make a token contribution to this without a strong UK involvement, and we suggest that it is most efficient to leave the resourcing of this to the groups who are driving this work.

Nevertheless the UK ALICE interest will benefit from in the long run from AliEN and other Alice Grid developments. This may be seen as a fair balancing against other areas which GridPP will be supporting.

ATLAS

The ATLAS document submitted describes a wish to develop the GANGA framework, installation kits, and an MC production system as short term vehicle to use Grid components. It includes a very extensive set of use cases and requirements which define well what is needed in the short term from EDG testbed components. ATLAs identify several common projects with LHCb.

The first project is an interface framework for using the GRID, called GANGA. GANGA will provide access to all GRID functionality ranging from metadata to physical file translation, job submission, job logging, and retrieval of results. There are clearly aspects of GANGA which are common to ATLAS and LHCb (integration of basic testbed-1 services) as well as experiment specific parts (eg only an experiment can define and interface to its metadata persistency , and provide policy for use of resources). This matches GridPP goals. This project is in common with LHCb.

The second project concerns an automatic application/environment specification and installation system. This is not primarily a Grid issue (in the sense that this would be done anyway in the absence of Grid), but is of course vital (as it is to all experiments)

for practical Grid operation. It is therefore a matter of PMB policy to determine GridPP resources to assign to this area. This project is common with LHCb.

The third project is a special use of GANGA for the Monte Carlo production system. This matches GridPP goals.

In all cases the programme programme presented is very comprehensive and evidently leads to a well defined adoption of EDG TB1 components. ATLAS and LHCb have worked together to produce this programme. The milestones are very detailed and are more than adequate for monitoring purposes.

LHCb

The LHCb document submitted prior to that of 30th Nov is an extremely clear and concrete document presenting their intent to build Grid applications on top of the EDG project, in parallel with existing software development, and where possible realise MDCs using these Grid applications. The document, although a draft still (and hence incomplete) lays out a clear schedule for use of the EDG testbeds. The final version of this document will be useful to GridPP for understanding and monitoring progress.

The further document submitted on 30th Nov presents very clearly defined projects and identifies common areas with ATLAS.

The first project is an interface framework for using the GRID, called GANGA. GANGA will provide access to all GRID functionality ranging from metadata to physical file translation, job submission, job logging, and retrieval of results. There are clearly aspects of GANGA which are common to ATLAS and LHCb (integration of basic testbed-1 services) as well as experiment specific parts (eg only an experiment can define and interface to its metadata persistency , and provide policy for use of resources). This matches GridPP goals. This project is in common with ATLAS.

The second is a project is a contribution to the further step to Gridify the complete Monte Carlo production chain using GANGA. This matches GridPP goals.

The second project concerns an automatic application/environment specification and installation system. This is not primarily a Grid issue (in the sense that this would be done anyway in the absence of Grid), but is of course vital (as it is to all experiments) for practical Grid operation. It is therefore a matter of PMB policy to determine GridPP resources to assign to this area. This project is common with ATLAS.

In all cases the programme programme presented is very comprehensive and evidently leads to a well defined adoption of EDG TB1 components. ATLAS and LHCb have worked together to produce this programme. The milestones are very detailed and are more than adequate for monitoring purposes.

CMS

CMS have submitted a document to the EB making clear that the CMS goals are tied to event production and analysis for MDCs using whatever tools exist, be they Grid like or not. CMS is more closely aligned with the CMS Grid work being undertaken in the US, and hence the PPDG project. We note that CMS have already integrated several Grid components into their processing, for example they already made use of GDMP for data management.

In this context CMS are not explicitly tied to making use of EDG testbed components to build their Grid applications, but will use whatever components exist from PPDG, DataGRID, Globus .etc. The TAG accepts that this is perfectly legitimate and follows a clear policy which makes sense given the strong US CMS effort in this area.

The CMS document submitted for 30 Nov requests posts in four areas

DBase support, admin and development. However goals are all general without concrete tasks and milestones. They suggests linking this to MDC goals. This would be difficult as MDC goals are in terms of number of events analysed, and as such can clearly be met without any Grid technology. This would perhaps fall more within the realm of support at the Tier-1 centre.

The second request is for a post to work on monitoring. This is much more narrowly scoped and hence clearly defined. The document presents some sensible first guess milestones, and openly suggests these can be refined.

MC/Analysis Grid framework: This is well described with clear and tangible milestones, and is similar in goals to that of most other experiments. This will require several FTE years effort in total, of which the UK would contribute a part. This matches GridPP goals.

Object level access: It is well recognised that object level access is a required feature which is not yet well addressed. This is certainly an area we would recommend GridPP support in collaboration with BaBar. Again, the milestones are good.

In summary, CMS has presented for the most part a set of well defined tasks against milestones. All but the first manifestly meet the overall goals of GridPP and provide tangible milestones. CMS UK staff will be working within an established and experienced Grid environment centred in the US.

4. Middleware areas requiring support / further development

WP3: The UK is responsible for WP3. As such the UK should support this at an appropriate level to allow the WP manager to execute their responsibilities. In addition to the EU funded effort a minimum of two people are required to underwrite a sensible resource base from which to lead this WG.

WP5: The UK is responsible for WP5. As such the UK should support this at an appropriate level to allow the WP manager to execute their responsibilities. In addition to the EU funded effort a minimum of two people are required to underwrite a sensible resource base from which to lead this WG.

5. Central Service Technical Requirements

There are a variety of roles for a central service during the lifetime of the GridPP project. These include

- A TierA centre in the BaBar distributed computing model providing a substantial service to the whole BaBar collaboration. This is required to reduce the UK's contribution to the BaBar Core Fund.
- A testbed in the DataGrid Project to meet the UK commitments to the sequence of testbed milestones
- A prototype Tier1 Centre in the LHC Computing Grid Project. The requirements for this are not yet known but the model of a Virtual Tier1 Centre goes beyond the EDG model.
- Part of the European end of the InterGrid and DataTAG projects involving CMS, Atlas, and BaBar
- Providing resources for existing and future experiments. Even the future experiments have data challenge plans which they wish to meet irrespective of the Grid.

It is hoped to achieve all of the above with a common infrastructure which can be split logically rather than according to projects. As equipment will be bought over the lifetime of the project, it is unlikely that it will all be the same so the plan should be to accommodate a variety of hardware within a common management infrastructure.

Requirements

Central services are expected to provide a higher availability and level of service than services run by University departments. There are a number of reasons for this:

a) The overlap of demands from a larger number of experiments means there are seldom slack periods in demand.

b) Dedicated staff with no competing physics demands have raised expectations over the years.

c) Central resources are usually used for experiment commitments rather than personal work so failure is more public.

In addition to the traditional demands from the UK community, the Grid has introduced several new ones

- A wider user community – users from outside the UK increases the support load.
- External constraints on the service – the EDG testbed is very prescriptive in software to be run even before its own software is installed. Meeting these requirements required system management effort.
- A variety of new services required in a grid environment have increased the complexity of the service (see below).
- The central service will act as conduit to the other UK sites from SLAC (for BaBar) and from CERN (for EDG). In both of these roles, the central service staff will support universities and provide training and expertise to backup the university staff.
- Development means widespread testing. A large service is a natural target for such tests. These invariably involve system managers too.

And some further increases in demand are foreseen:

EDG Testbed and InterGrid look like requiring a plethora of contacts for security, operations, networking, data which all raise the level of support needed for the central which participates in the projects above.

New Services

The following new services have already been identified

- Certificate Authority. The security infrastructure relies heavily on the integrity of the CA which signs certificates for UK PP users. There is a strict definition of tasks and a heavy layer of administration involved in delivering this service.
- Information Service: there are several of these required with hierarchies including a UK level which would be best provided at the Tier1 Centre which could provide a high level of availability through monitoring and on-call staff.
- TierA Centre: this service will greatly increase the number of users of the centre with a correspondingly higher level of support and over a longer time-frame to allow for time differences between UK and US.

Hardware Requirements.

The initial proposal is that the RAL centre will consist of

- Cluster(s) of PC running Linux under the control of a batch system (eg PBS) Dual cpus are not mandatory but currently give best value for money. Target of 256MB memory/cpu and scratch disk of 70GB internal to each box
- Standalone disk servers of the order of 1TB (this will increase with increasing disk size) running Linux. Having the same operating system as the cpu farm will enable i/o intensive services (event servers, skimmers) to run on the same machine as the disk. It is believed that a standard Linux disk server with IDE drives can meet all disk requirements for databases and flat files.
- An increase to the capacity of the recently refurbished STK robot by buying more tapes and an increase in bandwidth by adding tape drives. STK 9940 tapes hold 60GB and so give biggest capacity although 9840s (20GB) are faster and could be used if there is a need.

Existing equipment like the BaBar Suns and disk will be incorporated and should meet other requirements for Suns or Solaris. Eventually the existing CSF farms will be integrated too. Using estimates from summer 2001 we might expect an additional 200 cpus, 48 TB of disk, 50TB of tape and an additional 40MB/s tape bandwidth early in 2002 followed by a similar number of more powerful cpus and an increasing amount of storage in each of the two following years.

Staffing Estimates

The staff effort required has been calculated by breaking the foreseen work into specific areas. The FTE numbers do not necessarily map onto people in the categories shown; some numbers hide a blend of skills at different levels. As an example the environment at RAL has been used and the staff requirements reduced in some cases by sharing tasks or taking other sources of funding into account.

CPU Farms
System admin of the major farms (500 cpus this year and growing). All aspects of keeping the service running, patched, backed up. Installing new software to meet grid and user requirements. Batch software configuration. End-user support. Documentation (2.0 FTE)

Disk Servers
System admin of the disk servers (40 this year), maintenance of the filesystems, performance optimisation. (1.0FTE)

DataStore
Management and development of the integrated datastore system including robots, servers, local and 3rd party software, system admin of servers, user support and documentation. (2.5FTE, 0.5 from other sources)
Operations. Handling all manual interventions, physical import/export of tapes and other movements (eg backups into firesafes) (1.0 FTE)

Core Services
A collection of high availability services that, with the addition of corresponding grid services will grow to become a full-time job. AFS, /home filesystems, mail, NISS, software repositories, web servers, LDAP and other Grid services, (1.0 FTE)

Operations
Hardware interventions, replacement and repair, construction, physical assembly and installation for disk servers and cpu (1.0FTE)
System monitoring systems. Development and maintenance of the systems which will log problems and alert the other system managers (0.3FTE)
Installation Management – dealing with all the issues of racking, power provision, safety, cooling, delivery (0.4FTE)
Helpdesk – single point of phone contact, running helpdesk service. (1.5FTE – 1.2 from other sources)

Registration/CA
Technical and management issues of developing and running a Certificate Authority and interfacing with other bodies (1.0 FTE – shared equally with e-Science Centre)
Operation of CA – issuing certificates, revoking them, security and integrity of service. (0.6 FTE - shared equally with e-Science Centre)
Registration of traditional userids. It may be that this will drop off in future but no sign yet and may grow greatly for BaBar. If grid userids become dynamic there will be an auditing load. (0.3 FTE)
Accounting, reports, statistics (0.2FTE)

R&D
A distinct cluster of testbed machines is required for testing new things and working with other sites who wish to do tests with RAL as the remote end. Hardware costs are neglible as cast-off machines will usually suffice but staff effort is required to administer them, and the collaborative effort, negligible for each request, accumulates to a noticable total (0.5FTE)
Keeping up with the latest developments and gaining experience with new technology requires effort. With continuous installation of new equipment the effort required to judge the relevant cpu, disk, tape, and network solutions is substantial. (0.5FTE)

Networking
National monitoring and other international work is included here (0.4FTE)
Tier1 will put a load on the RAL networking infrastructure. There is (a) infrastructure work on: JANET connections, firewalling, routing, DNS, installation, cabling, and trouble-shooting and (b) local management of the HEP resources (0.4FTE)

Testbed
Administration and development of DataGrid testbed, participation in testbed rollouts, local troubleshooting of datagrid software. (1FTE – funded by GridPP for WP6)

Tier2 liaison
Rolling out the datagrid to other UK centres. Similar work to WP6 post but directed to other UK centres (1.0FTE)

Other HEP Services
BaBar and CDF have dedicated hardware at RAL. Running and maintaining this requires a whole range of duties similar to most of the above. (0.5FTE, 0.2 from CDF)

Management
Management and planning of all the above tasks and staff (1.0 FTE).

Summary

	FTE	TierA/1
CPU Farms	2.0	2.0
Disk Servers	1.0	1.0
DataStore	3.5	3.0
Core Services	1.0	1.0
Operations	2.0	2.0
Registration/CA	2.1	1.3

R&D	1.0	1.0
Networking	0.8	0.8
Testbed	1.0	0.0
Tier2 liaison	1.0	1.0
Other HEP Services	0.5	0.3
Management	1.0	1.0
	16.9	14.4

6. Tier-2

There is some ambiguity concerning the size of tier 2 sites, so a site consisting of some 1000 cpus and around 75 TB of RAID storage is considered. Other assumptions include that the site has no mass (tape) storage, that software and OS upgrades are performed in a highly automated manner and that individual experiments provide installations kits in the form of rpms or similar. The management of these sites will consist of:

- Installation
- Network management
- Repair of hardware failures (with equipment suppliers)
- Asset tracking
- Software installations and upgrades
- User support
- Essential health and safety (e.g. electrical safety testing)

These activities requires access to a diverse skill set which must be available at any site hoping to be a Tier 2 site. On average the management of a tier 2 site will require the equivalent of 2 ftes, however, unlike the Tier 1 site, the workload will not be flat but will have periods intense activity (e.g. electrical safety testing) and periods when only routine maintenance is required. Management of a mass storage device would require the equivalent of an additional fte.

7. Glossary

EDG: European DataGRID project

EGG-TB1: Testbed release 1 of the EDG.