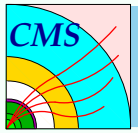


CCS

CERN, Tuesday 2 Nov 2004

WorkLoad Management

Stefano Lacaprara

Stefano.Lacaprara@pd.infn.it

INFN and Padova University

Outline

- News,
- Draft milestones and dates,
- People and activity in various area,

Draft milestones with dates

- After gridPP (UK) and Lucas request, tentative “agenda” for next year for Workload Management objective
- Try to define areas of work (already advanced) tasks
- Try to define what should be achieved in what time
- Should match as much as possible CMS/CCS milestones (not yet synchronized...)
- Try to identify people/institute for each task/sub task
- Find eventual man power shortage
- Much of the work for Workload management depends on other’s workplan

WM dependencies

- Data management
  - Access to data depends a lot on how data will be distributed
  - Different pattern leads to different distributed scenarios
- Infrastructure deployment
  - LCG/Grid3/NorduGrid
  - gLite EGEE deplyoment
- Analysis model
  - Still not fully defined for CMS
  - More or less ok for next few months
  - Will stay like this until PTDR?
  - Will change considerably afterward?
- Working scenario: much depends on other
- Hard to look very far in future
- So, define deliverables and timescale until end 2005 (PTDR)

WM Areas of activity

- Data Publication
- Access to resources
- CMS software deployment
- Job preparation
- Job splitting
- Monitoring and bookkeeping
- Output retrieval, storage and publication
- Training and documentation

High level deliverables

- End 2004 Prototypes for high level tool for WM and simple use cases
- 03/2005 Extension of usage of prototypes to a wide audience of PRS members to access remote data w/o data movement
- 06/2005 Evaluation of new functionality of LCG projects (EGEE) for job submission and analysis framework
- 09/2005 Extension to cope with more complex data distribution scenarios, including data movement on demand
- 09/2005 Effective, wide usage by PRS for PTDR studies
- 03/2006(??) Ready for DC06: effective, quasi on-line, realistic analysis during DC06 data challenge
- ...
- ??/2007(?) Ready for data taking

Lower level deliverables

- More precision of WM deliverables wrt to the high level one (pretty general)
- To be reviewed, re-discussed and eventually modified or redefined as the project unfolds
- Dependency on other projects, notable **Data Management**, is very strict
- APROM discussion to coordinate timescale, requirements, etc...
- Will present deliverables and timescale for each areas defined before
- Will also discuss actual status and future plan

Data publication

- End 2004 Definition of requirements for publications. What, where and how. Prototype usage for job preparation/splitting.
- 03/2005 Evaluation of LCG dataset catalogs for CMS publication.
- Implication of DataManagement prototype for Publication schema: synchronization, duplication of info, etc...
- Publication of private or group-wide data.
- 09/2005 Deployment and testing for data published for PTDR, including complex schema, such as distributed dataset, streams, skims, etc

People

- Production people: Alexei, Julia, Tony, ...
- WM: Alessandra will focus on that
- Requirements, input, discussion also from WM tool developing groups (grape)
- Need input also from Data Management (discussed during DM workshop)
- Probably enough people
- Need more involvement by APROM (integration with DM)

Status

- Very active discussion about requirements
- What should go into CMS publication system
- and what is grid responsibility
- at which level should we integrate with DM
- Problem in achieving a effective discussion! (common to many CMS groups!)
  - Difficult by mail exchange (achieved rate 1 mail/5 min– a MailChallenge?)
  - Difficult by “standard” meeting with presentation, even useful to drive discussion
  - This week (tomorrow?) technical discussion about scope and roles for PubDB

Access to resources

- End 2004 Usage of available testbed on various T1, eventually T2
- 03/2005 Deployment of RB able to interface to CMS dataset catalogs
- Deployment and initial test of gLite testbed
- 09/2005 Wide usage of remote resources by PRS users.

People:

- T1 center people . . .
- Need to define better responsabilities, but actual sistuation is already $\sim$ fine
- Need to integrate with grid deployment: drive test bed for our needs

Big issue:

- How to deal with priority?
- Today only possible at CE level, on the hand of local site manager (if LBS allows)
- Already today some delay in testing due to farm usage by non-CMS experiment!

Software distribution

- End 2004 Definition of requirements for sw deployment and tag publication
- 03/2005 Test of Sw deployment on GRID resources
- 09/2005 Prototype for (semi-)automatic sw deployment world-wide

People

- Nick, Claudio, Karlsruhe, DAR-group
- Well covered – maybe too much :)
- Nice start with Claudio documents on requirements, need to move further on integration of different existing tools
- GAG software installation document at `http://cern.ch/fca/DCFeedBack.doc|pdf`

Job preparation

- End 2004 Prototype working. Simple use case, including private sw and executables. Access to published information
- 03/2005 Enhanced prototype
  - more complex configuration as defined by user feedback on simple case.
  - Training of PRS community.
  - Access to Dataset catalogs by LCG RB.
- 09/2005 Full working tool

People

- Grape, Gross for LCG
- RunJob for UAF$\rightarrow$MOP$\rightarrow$LCG
- Mario Kadastik on NorduGrid: first experiences
- Julia for EGEE (need more!)
- Well covered, need more integration
- This week, meet with David Collins for Grape/Gross common developing and integration: focused on LCG

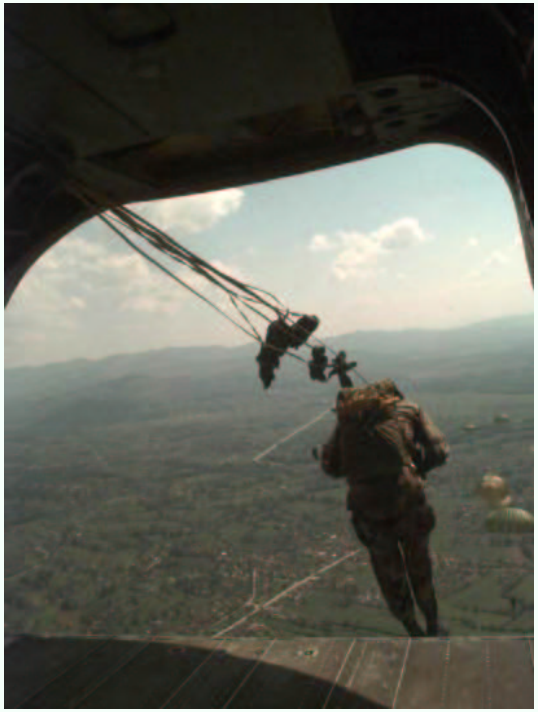
Job preparation

Status

- Progressing, bit slower than expected!
- Last week first successful jobs submission with grape *using* PubDB as dataset catalog (first time!)
- Progress slowed by many problems found (and being addressed)
- Biggest was interface with PubDB
- PubDB is evolving, need stable interface
- Only hitting real problems we are understanding what we need from PubDB (Not a surprise!)
- Advantage of approach: *start developing something simple and try to make it works. Then learn from it*
- Good status for RB$\rightarrow$PubDB interface by Heintz and Flavia: prototype ready, asked for dedicated deployment of RB for test

Job preparation (II)

- Two experience on Grid3 (Rick) and NorduGrid (Mario)
- Very interesting results!
- Approach is to use “opportunistic” resources: *paratrooper* approach



- Carry with you all you need
  - Install sw privately
  - Move chunk of data on remote resources
  - Run on it
- Different wrt what I presented as WM workplan
- GOOD! We do need to play around with different scenarios!

Job preparation (III)

- Is this approach a choice or a forced solution given the Grid3/NorduGrid architecture?
- Risk: need too much resources to use resources!
- Install al CMS sw: very expensive! Then move data (cost depends on data size, if small no problem)
- Run on data: if running is fast, overhead can be impressive!
- Can be improved if cluster-job is done
- Mother job install sw, sons use it
- Requirements from this experience: tool to split data in small independent chunks (even smaller than a single run)
- Do we need this?

Job preparation (IV)

- Much depends on data/event model
- Actual model assumes dataset complete on a Tn
- Must evaluate carefully priority. I don't think that this is a first order priority.
- First need to move datasets with appropriate movement publication
- Phedex provide functionality: need to integrate with publication schema
- Then move on fancier data movement scenarios

Job preparation (V)

- Related issue
- Do we need dataset splitting for job splitting? (Also EGEE approach)
- Not mandatory!
- Complete Dataset (or big part of it) can be on a SE, and accessed by many splitted jobs from WNs of close CE
- If data is splitted, then the job splitting *will* follow more strictly data splitting
- If, for better resource usage (eg very fast job reading many runs), need to re-join small pieces of dataset. Can be complex!
- Not guaranteed that best use of resources is done.
- Need to ensure that balancing is done
- Implication for data and event model

Job splitting

- End 2004 Simple splitting based on user configuration, assuming un-splitted dataset
- 03/2005
  - More complex splitting assuming also partial datasets distribution among different site.
  - Evaluation of new LCG functionality for job splitting.
  - Prototype for job clusters submission.
- 09/2005
  - Integration of CMS specific knowledge in job splitting with resource matching done by LCG
  - Splitting done according to data and resource distribution

People

- Same as job preparation
- Not yet idea about how to do a real job clustering matching data *and* resources in an optimal way

Monitoring and bookkeeping

- End 2004 Very simple prototype working. Definition of requirements for application monitoring and bookkeeping.
- 03/2005 Evaluation and performance analysis of existing tools (BOSS, MonaLisa, JAM,...). Integration with job submission for resubmission on fault, detection of black holes, etc...
- 09/2005 GUI/WEB front-end for physicist. ...

People

- MonaLisa, BOSS, JAM
- Well covered –too much? :) for basic services
- Still low experience on what user really want to monitor and bookkeep

Output retrieval and publication

- End 2004 User get back its output easily.
- 03/2005 User can also publish output on grid storage resources with coherent publication
- 09/2005 as required by users...

People

- Same people as job preparation
- Limited experience on publication
- No experience, and limited ideas for publication

GAG: software installation document

Available at
http://cern.ch/fca/DCFeedBack.doc
http://cern.ch/fca/DCFeedBack.pdf

- We are requested to give feedback/approval/etc...
- Feedback already given in the past, many changes went into document
- Goal: we want to be able to do in future what we are doing now
- If service provided to do better job, fine
- Document is quite general (obviously), we should check if this is fine with our view
- IMHO is (now) fine, but I'd like to have also other people looking at it