Tier-2 - Analysis, Groups, Data, ...
A Collection of
Status & Problems & Needs

PADA Meeting

CMS Plenary Week, CERN

December 10th, 2008

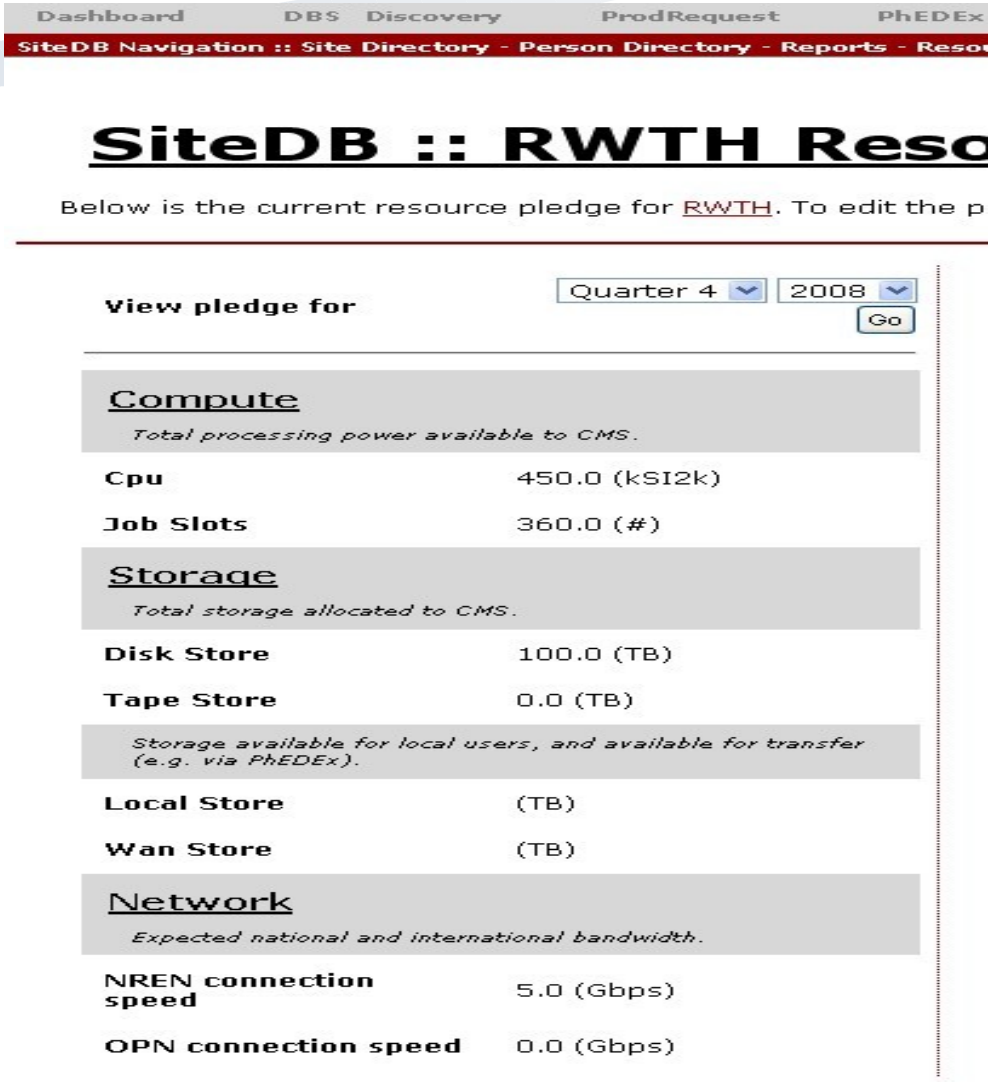
More general presentation about Analysis @ Tier-2s
in next Friday's
Offline/Computing Plenary Meeting

Thomas Kress, III. Institute B

RWTH AACHEN UNIVERSITY

Resource Information & Planning

- SiteDB is a nice tool, we could take more benefit out of it for the resource planning
- Pledge information is not always consistent and not really up-to-date for all sites
- Proposal:
  - Use current status instead of / (in addition to) pledges (pledges are, at least on a yearly basis, documented in WLCG tables)
  - Re-consider and simplify/adapt categories, define keywords exactly, since some seem to be confusing for some sites
    - Which level of granularity is necessary / sensible ?
  - Ask T0-T2 sites to keep status of current quarter up-to-date
  - Give also estimates for following two quarters



Summer08 Data Distribution

- We (DataOps & ThK) are in the process of distributing RECO and AODsim data to the T-2s centrally
  - Only samples requested by physics groups, but not the detector requests (ID=0)
  - AODsim w/o buggy Madgraph samples
  - One site in EU, another outside of EU
- Good progress !
  - A good fraction of the samples was already transferred by groups / T2 sites
  - Used a vast number of sites; group locations partly taken into account, although not required in the distribution model
  - Official „central data“ sites for a dataset are marked in blue; might extent this to yellow for data coounted towards group space
  - So far no automatic accounting of central or group space possible

ID	Dataset	# requested events	# events	% done	Total size [TiB]	T1 for custodial storage	T2 (% of dataset available)
4003040	/BBJets1000toInf-madgraph/Summer08_IDEAL_V9_v1/GEN-SIM-RECO	300000	0	0.00	0.000	RAL	
4030010	/BBJets100to250-madgraph/Summer08_IDEAL_V9_v1/GEN-SIM-RECO	1000000	0	0.00	0.000	CNAF	
4003020	/BBJets250to500-madgraph/Summer08_IDEAL_V9_v1/GEN-SIM-RECO	1000000	0	0.00	0.000	RAL	
4003030	/BBJets500to1000-madgraph/Summer08_IDEAL_V9_v1/GEN-SIM-RECO	10000000	0	0.00	0.000	IN2P3	
1038000	/BtoJpsiMuMu/Summer08_IDEAL_V9_v2/GEN-SIM-RECO	2000000	2434076	100.00	0.628	FZK	T2_CH_CSCS (32.0) T2_CN_Beijing (100.0)
0	/CosmicMCBOff10GeV/Summer08_COSMMC_21X_v4/RECO	30000000	30131500	100.00	10.280	CNAF	
0	/CosmicMCBOff4to10GeV/Summer08_COSMMC_21X_v3/RECO	20000000	20685000	100.00	6.942	FZK	T2_US_Caltech (100.0)
0	/CosmicMCBOn10GeV/Summer08_COSMMC_21X_v4/RECO	30000000	30310000	100.00	10.345	FNAL	T2_US_Florida (93.3) T2_DE_RWTH (2.3) T2_IT_Legnaro (91.0) T2_IT_Pisa (91.0) T2_US_Purdue (91.0)
0	/CosmicMCBOn4to10GeV/Summer08_COSMMC_21X_v4/RECO	20000000	19254000	96.27	6.473	FNAL	T2_US_Florida (3.0) T2_DE_RWTH (2.0) T2_IT_Legnaro (90.3) T2_IT_Pisa (90.3)
1061030	/DYmumuM1000/Summer08_IDEAL_V9_v1/GEN-SIM-RECO	10000	10296	100.00	0.003	CNAF	T2_ES_CIEMAT (100.0) T2_US_UCSD (100.0) T2_US_Wisconsin (100.0) T2_DE_RWTH (100.0) T2_FR_CCIN2P3 (100.0) T2_IT_Legnaro (100.0)
1061010	/DYmumuM200/Summer08_IDEAL_V9_v1/GEN-SIM-RECO	10000	10022	100.00	0.003	CNAF	T2_ES_CIEMAT (100.0) T2_US_UCSD (100.0) T2_US_Wisconsin (100.0) T2_DE_RWTH (100.0) T2_FR_CCIN2P3 (100.0) T2_IT_Legnaro (100.0)
1061040	/DYmumuM2000/Summer08_IDEAL_V9_v1/GEN-SIM-RECO	10000	10091	100.00	0.003	CNAF	T2_ES_CIEMAT (100.0) T2_US_UCSD (100.0) T2_US_Wisconsin (100.0) T2_DE_RWTH (100.0) T2_FR_CCIN2P3 (100.0) T2_IT_Legnaro (100.0)

Tier-2 Data Distribution - con't

- Distribution of data to Tier-2 sites will become an everyday business
- Although great support without latency by DataOps (Guillelmo, Sie, Maarten), in my opinion development towards more automation is needed
- From today on, we have a possibility for an accounting for central & group data
  - For new data transfers only !
  - CSA07 and other older data will be more or less phased-out and deleted/“made group-/private“ soon
  - CRAFT re-reprocessing and new FastSim production with new Phedex accounting
  - Summer08 MC sets will be used extensively during the next months
    - Any change that we could also inject the Summer08 MCs, already distributed to the „official“ Tier-2 sites, into the new Phedex ? Otherwise a lot of handwork necessary to have a complete accounting of central and group data at the Tier-2 sites
    - Not necessary if all Summer08 data will be re-reprocessed and re-transferred

Group / Tier-2 Associations

- It took us quite some time to consolidate the year 2008 associations
- Detector & physics group hosting at Tier-2 now relevant for MoA credits
- We promised to regularily review and extent these associations
  - Tier-2 utilization by groups just started, so no sense to review choosen choices already now
  - At least two established sites (BR_UERJ, BE_UCL) are now ready to support another / a first group
    - Forward Physics and Tracker - not the „hottest“ wishes , so desired allocations IMHO seem to be OK
  - Have to decide from when to allow adding new associations officially (Jan 1st ?) and on general policies

User Home Storage Space

- We anticipate the need of 0.5 – 1 TB per user as stage-out area for 2008/9
  - Some sites will add additional national/local resources transparently
- Most countries operate Tier-2 site(s) to support their national users
- In case a Tier-3 with sufficient CMS Grid cabability is available, do we allow to have an official user area there ?
  - Im principle yes, but this implies the question of support policy for private Tier-3 sites
- Users with a CERN affiliation can be provided with home space at the T2 part of the CAF
  - How to define „CERN affiliation“ ? (e.g. CERN authorship on publications/notes)
  - About 50-100 users
- For countries w/o a Tier-2 (or GRID Tier-3) we have to try to find „Tier-2 friends“
  - I guess about 5% of our users are affected, a list of affected countries exists
  - A good deal would be: „we provide you with user storage, you participate in the Tier-2 support operation“
  - The fall-back CERN CAF-T2 is not really appriciated because of a lack of resources there

User Home Storage Space - con't

- LFN of the users' home dir. is */store/user/<HN-name>*
  - Will be mapped to a (transparent) physical dir. by a site's rule
  - We will have one single entry User ↔ T2 location stored in SiteDB
    - Status ?
  - CRAB will use this entry as the transparent default for a stage-out (if sandbox for small output is not used) of the job output, but the user can overwrite with any Tier2-3-.../SE name and path (but needs a local mapping and write-permission)
- Data in /store/user/<a_user> can be registered in local-scope DBS
  - Should be specified before the CRAB job is submitted
  - Can be a local-scope instance at CERN or at a Tier-2 which has installed a DBS system
    - To be discussed whether T2 DBS systems are sensibe and who (e.g. associated groups & locals) should/could/may use it
  - If a dataset is registered, other users can access it by Grid jobs

User Home Storage Space - con't

- Popular use case: a group's representative runs CRAB sub-skims on group's data
  - Per default the output is stored in his/her private storage area, which is usually disliked
    - Probably possible to use SE name/path overwrite option and stage-out to special area at group's Tier-2, but how to discriminate between private and group mapping of user's certificate (VOMS role possible ?) ?
  - The group likes to replicate the CRAB produced sub-skims to the other group's Tier-2 sites
    - To run another CRAB sub-skim with different stage-out location is not really elegant
- In any case an injection from the (semi-)private into the official storage area is urgently needed and was promised to the groups
  - To allow Phedex transfers/deletion and a full official data tracking
  - The idea is to transfer (with some restrictions) files into a /store/results/<group> area, and to inject into Phedex and global DBS
    - The tool is not yet there, and it seems there is no manpower for a rapid development
  - Serious problem ! When will such a tool be avaialble and what are possible/feasible workarounds ?

Storage Quota Tracking

- /store/data/ and /store/mc
  - (In my understanding) Phedex keeps LFN structure in transfers
    - Means, that same storage directory structure for central and group data if data transferred by Phedex
    - Unclear to me, how to handle future /store/results/ ?
  - From today on, DPG/POG/PAG & FacOps & DataOps markers become available – Great !
    - Accounting and querying of central vs. group data possible on DB level
    - But only for new tranfers ! Could we somehow bring the Summer08 data into the new accounting system ?
- /store/results/<group>
  - To be accounted towards group quota, details still under discussion, not yet implemented
- /store/user
  - Since at most only partial registration in local-scope DB, no non-ambigious central information about official user quota
  - Some users will have additional national/local resources added transparently
  - Some users will have storage resources at more than one site
  - Local T2 site probably has to monitor /store/user/* and /store/user to protect storage against overquota, since for most setups the Grid storage systems do not offer quotas
    - With all this restrictions, does is makes sense / is it necessary to have user quota information centrally ?

CPU Batch Fair Share & VOMS

- So far we have /cms and some /cms/<national> VOMS groups and some VOMS roles for e.g. production, …
- (How) do we have to extent this for the groups ?
  - If more or less everybody would register for every group, no practical benefit and only large overhead for group's administration
  - Roles for CRAB production of e.g. higgs secondary skims probably more useful
- Not trivial to adjust and monitor fair share values for large number of VOs, VOMS groups, VOMS roles, local submission, …
- If we want to use it, we have to define how - now !

Group / User Support & Documentation

- Groups (data) representative(s) should act as link person to the Tier-2 operation support team
  - For group data, the group representative(s) make the Phedex requests, local T2 support accepts/rejects
- How to officially inform the associated groups about Tier-2 operation problems ?
  - Tier-2 sites announce downtimes in official Grid DB, and in parallel send Emails to CMS computing HN
  - Just do a CC to the group's HN mailing list ?
  - In case of longer (several days) scheduled downtimes, replication of group's data to other Tier-2s nec. ?
- How get users help with problems related to Tier-2 sites ?
  - For home space usually direct national/local contact and „natural“ support possible
  - Are „group users“ allowed to contact Tier-2 site ?
    - Directly or via group contacts ?
    - Savannah, GGUS, ..., Hypernews, individual Emails, ...? Needs a decision soon !
- Unfortunately documentation is always/most of the time the last bullet ...
  - Me to be blamed !

Executive Summary

- Several things are in progress
- Not all is setup completely or statisfactorily yet
- Some decisions have to be made soon
- Manpower urgently needed for development of missing tools and/or intermediate workarounds