PD2P for Tier 1 Implementation Plan

Kaushik De

Univ. of Texas at Arlington

SW&C Workshop, CERN

Apr 6, 2011

Overview

- PD2P for Tier 2's working well
- Need PD2P for Tier 1's
  - Discussed in Naples Retreat as essential for the large data volumes expected during 2011-12
  - Especially important if we have 400 Hz data rates
  - Necessary for resource model described by Jim on Monday
- Start with simple implementation
- Evolve with experience

T1 -> T1 PD2P Proposal

- Primary copies of dataset are pre-placed at Tier 1's
  - See talk by Ueda on Monday
- Secondary copies at T1 by PD2P based on usage
  - Panda keeps track of the number of times a dataset subscribed by PD2P is used by user analysis jobs - nused
  - PD2P will make secondary copies = log10(nused)
  - So far, highest value of nused ~1000
- Location of secondary copies based on MoU share
  - Secondary copy will be at a different Tier 1
  - Use MoU share to choose replica location
- Dataset containers will be split
  - Datasets within a container will be split among Tier 1 sites, but according to MoU share

Additional Copies at Tier 2

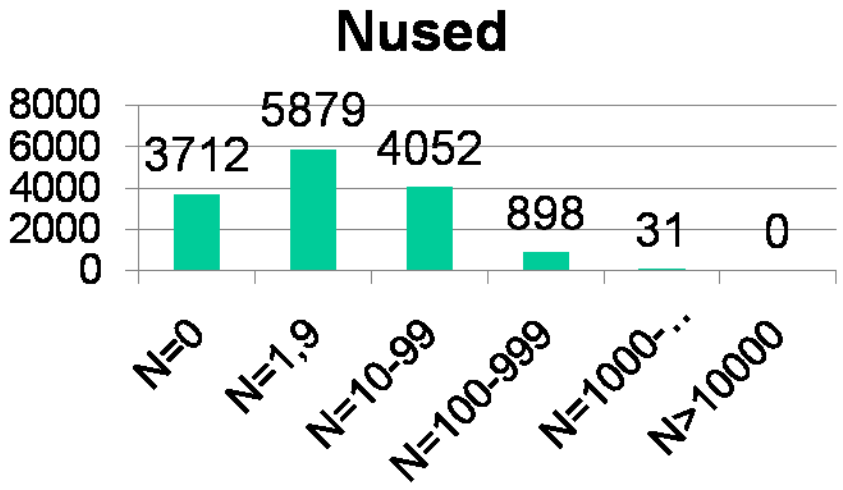
- Currently, only1 copy is made by PD2P at Tier 2
  - First copy is always made on first use of dataset
  - Additional copies only if site offline, subscription failed…
- Additional sites will now be added based on job backlog
  - If many jobsets are waiting to run
  - and if large number of jobs are waiting to run
  - add more sites based on log10(weight*nWaitingJobs)
  - The thresholds will be tuned based on experience
- Minimize sending additional copies to Tier 2 in same cloud
  - Negative weight if cloud already has Tier 2 copy
- Put some protection for maximum number of copies
  - Check always maximum 'complete' dataset copies < 5

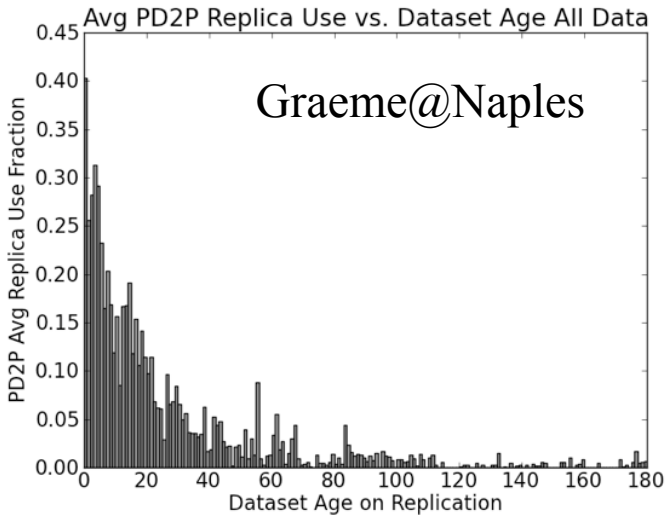
What about Tier 2D's

- The proposed T1 PD2P model still follows cloud boundaries
  - This is not necessary, but avoids transfer problems
- Once Tier 2D's are fully tested and configured
  - Subscriptions can be made without specifying source, so DDM topology will be used, simplifying data transfer to Tier 2D's
  - Effectively, this will break all cloud boundaries for PD2P
- Maybe Tier 2D=Tier 1?
  - This requires some additional development – we will evaluate if this is necessary after few months
  - Anyway, we may have to tune thresholds to get correct number of copies world-wide

Further Improvements

- If number of copies based on Panda usage does not work
  - We may look at DQ2 popularity and other metrics
- If number of copies based on waiting jobs does not work
  - We may add new table in PandaDB to track history
- If first copy is too expensive
  - We may add age of dataset, as Graeme showed at Naples



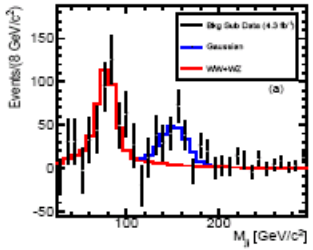


Replication Based on Waiting Jobs

- Snapshot of waiting jobs from this afternoon
  - 54139 - data10_7TeV.periodD.physics_L1Calo.PhysCont.ESD…
  - 11298 - mc10_7TeV.208802.Hijing_PbPb_2p75TeV_MinBias_Flow_JJ.digit.RDO…
  - 8393 - data10_7TeV.periodF.physics_JetTauEtmiss.PhysCont.ESD…
  - 6464 - data10_7TeV.periodI.physics_JetTauEtmiss.PhysCont.ESD…
  - 4602 - user.koffas.data10_7TeV.periodGHI.physics_JetTauEtmiss…
  - 3974 - user09.AlbertoAnnovi.FtkV14_base_20091204_130347
  - 3815 - data10_7TeV.periodE.physics_JetTauEtmiss.PhysCont.ESD…
  - 3608 - data10_7TeV.periodH.physics_JetTauEtmiss.PhysCont.ESD…
  - 2914 - data11_7TeV.00178109.physics_JetTauEtmiss.recon.ESD…
  - 2615 - mc09_7TeV.108495.PythiaB_bbmu4X.recon.ESD…
  - 2502 - data11_7TeV.00178044.physics_JetTauEtmiss.recon.ESD…
  - 2187 - data10_7TeV.periodB.physics_L1Calo.PhysCont.ESD…
  - 2168 - data10_7TeV.periodC.physics_L1Calo.PhysCont.ESD…
- With proposed algorithm, the above datasets could have additional copies (caveat: I did not check jobset threshold for all – top of list passed)
- But based on policy (no ESD, RDO, user), these would not be replicated
- Hopefully, we will have AOD, DESD, NTUP etc waiting in the future

Conclusion

- If proposal is accepted today, we can start T1 PD2P
- Could be ready in a few days
- Future tuning will be based on experience



http://arxiv.org/PS_cache/arxiv/pdf/1104/1104.0699v1.pdf

BACKUP SLIDES

Current Policies

- General policies:
  - Only T1-> T2
  - Only within cloud
  - Special algorithm for EOS (CERN)
  - Only for project names mc*, data*
  - No RAW, HITS, RDO and ESD replication allowed
  - Currently skipping Group Production data
  - Skip GangaRobot, HammerCloud, _dis, _sub etc
- Cloud selection:
  - Cloud T1 must have data to be replicated
  - Cloud must have at least 1 site with PD2P enabled (better >1)
- Site selection:
  - Site must have free space
  - Site must be analysis, not test, online…