



PD2P Changes for Summer

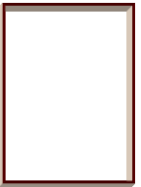
Kaushik De

Univ. of Texas at Arlington

ADC Weekly Meeting

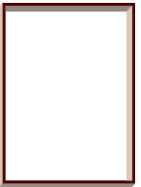
July 4, 2011

Background



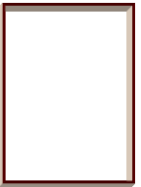
- New PD2P algorithm in use since April 18th
- Includes both Tier 1 and Tier 2 sites
 - Tier 1 distribution based on dataset usage
 - Tier 2 distribution minimizes job wait time
- Algorithm description:
<https://twiki.cern.ch/twiki/bin/viewauth/Atlas/PandaDynamicDataPlacement>
- First results shown at ADC TIM meeting in Dubna
 - <https://indico.cern.ch/materialDisplay.py?contribId=4&sessionId=5&materialId=slides&confId=132486>
 - In general, new PD2P working well
 - No complaints from users
 - But, amount of data distributed, especially for Tier 2 sites, very low
 - No more pre-placed data at Tier 2's (only PD2P and DaTri)
 - ESD and RAW data distribution not allowed for Tier 2's

Proposed Plan



- Many discussions during past few weeks
- Consensus goals (for summer):
 - we try to increase data distribution to Tier 2's
 - evaluate results after two months
- Plan of action:
 - Simone's idea: make extra copy using 'fair share' whenever a first copy is made at Tier 2 by PD2P
 - Ueda's idea: make additional 'fair share' Tier 2 copy every time there is a copy at Tier 1
 - Alexei's idea: make 2 pre-placed 'fair share' copies at Tier 2 of all group produced NTUPLE datasets
 - Simone's suggestion: DQ2 concept of 'closeness' should be used along with brokerage, to improve subscription performance
 - ADC will define grouping of sites (based on capability) for 'fair share' placement of data at Tier 2's

Actions



- This plan was approved by CREM last week
- Will be implemented in a few days
- Same 'fair share' tables will be used for pre-placed and PD2P data distributions
- We will have a lot more knobs (and flexibility)
- We will have a lot more copies – at least 2-4 times more data at Tier 2's