



# **Tier 3 Data Management Proposal**

**Kaushik De**

**University of Texas At Arlington**

**US ATLAS Tier 2/Tier 3 Workshop, UC**

**August 20, 2009**

# Look at some Use Cases



- ❑ Use Case 1:
  - ❑ Run pathena jobs at Tier 2, bring output back to Tier 3 for viewing
  - ❑ Easy – but need robust dq2-\* client services
- ❑ Use Case 2:
  - ❑ Run pathena jobs at Tier 2, create DnPD, asynchronously route output back to Tier 3 (delete from Tier 2, catalog at Tier 3)
  - ❑ New feature request from pathena
- ❑ Use Case 3:
  - ❑ Request specific DnPD dataset to be replicated to a Tier 3 site from a Tier 1 or Tier 2 site (within same cloud)
  - ❑ Can we do this without full DQ2 subscription mechanism at Tier 3?
- ❑ Use case 4:
  - ❑ Run pathena job at Tier 3, input and output both reside at Tier 3
  - ❑ Easy case – set up site as Tier 2'

# Some Caveats



- ❑ Tier 3's running production (like UTD, Wisc and Illinois)
  - ❑ These Tier 3's are configured like Tier 2's
  - ❑ Analysis use case can be similar to Tier 2 – special use cases not too relevant for them
- ❑ Only considering pathena based analysis
  - ❑ Probably safe assumption in the US, but may not work ATLAS-wide
- ❑ There are many shades of grey
  - ❑ Many users will invent their own analysis model
  - ❑ Many sites will invent in-between use cases
  - ❑ Need to be flexible and evolve
  - ❑ But also need some rigidity since support effort is limited

# But what is a Tier 3?



- ❑ We know there are many types of Tier 3's
  - ❑ Do all use use cases apply to all types?
  - ❑ Let's look at the classifications from Tier 3 task force
- ❑ Tier 3 gs
  - ❑ Center with full grid services – Tier 2' (like U. Wisc.)
- ❑ Tier 3 g
  - ❑ Cluster with grid connectivity
  - ❑ Probably the most common type of Tier 3 – focus on this
- ❑ Tier 3 w
  - ❑ Individual workstation – RootTuple analysis, grid submission
  - ❑ Limited data management capability – dq2-\* client tools sufficient
- ❑ Tier 3 af
  - ❑ Most difficult use case – not well defined yet

# Mapping Use Cases to Tier 3 Types



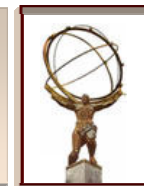
	T3 gs	T3 g	T3 w	T3 af
Use Case 1: Get pathena output			✓	✓
Use Case 2: Bring pathena output to T3 for shared analysis	✓	✓		✓
Use Case 3: Subscriptions to Tier 3	✓	✓		✓
Use Case 4: Run pathena at Tier 3	✓			✓

# Technical Implementation – Use Case 2



- ❑ Discussed with Torre, Alexei, Jim C., Doug & Michael
- ❑ Run pathena jobs at Tier 2, create DnPD, automatically and reliably transfer output back to Tier 3 (delete from Tier 2, catalog at Tier 3)
  - ❑ Propose new `--destination` option in pathena
  - ❑ Use pandaMover to transfer data back to Tier 3
  - ❑ All Tier 3's with this option must implement SRM v2.2 or greater
  - ❑ Single LFC maintained by BNL (pathena will restrict `--destination` option to this LFC only)
  - ❑ Data will be shareable ATLAS-wide (not only within Tier 3, though this is the most common use case)
  - ❑ Will need automatic data consistency and data cleanup tools

# Technical Implementation – Use Case 3



- ❑ Request specific DnPD dataset to be moved to a Tier 3 site from a Tier 1 or Tier 2 site (within same cloud)
  - ❑ Option 1:
    - Use pandaMover
    - Will require some development by Panda team
  - ❑ Option 2:
    - Use DQ2 site services
    - Single SS machine maintained by BNL
  - ❑ Option 3:
    - Use BitTorrent
    - Needs new development (and resolution of security issues)
- ❑ Some open questions:
  - ❑ Restrict data movement from Tier 3 sources (new ToA flag)?