

# OSG DDM Operational Tools

Patrick McGuigan

January 26, 2007

WLCG Collaboration DDM  
Workshop

# Current Tools

- Scripts that we use are at:
  - <http://atlas-sw.cern.ch/cgi-bin/viewcvs-atlas.cgi/offline/Production/panda/scripts/>
- `cleanse.py`
  - Purpose is to remove files from DQ2 storage location on Tier 2's
- `checkse.py`
  - Determines state of files found in DQ2 storage.

# checkse.py

- Examines all files below a specified local path
- LFN is taken from basename of file
- Determines if the file is a candidate for removal based on:
  - Existence at Tier1
  - Existence in LRC
  - Age of file
  - Archival Flag
- When removing file
  - Delete entries in LRC
  - Delete file from disk

# cleanse.py (cont.)

- Practical results
  - Script is slow
    - At UTA, process about 10K files per hour
    - Query made to Tier1 for each file
    - Can fix with bulk lookups
  - Never deletes orphaned files
    - Files on disk but not in LRC
    - Misidentified as too new
  - Can create orphans
    - Executor may not have deletion privilege on disk file
    - Multiple PFN's for single guid

# checkse.py

- Based on cleanse.py
- Finds problematic files
  - Files not in LRC
  - Files not at Tier1
- Output is structured better than cleanse.py
- Suffers from the slowness seen with cleanse.py