



# **PanDA in a Federated Storage World**

**Kaushik De**  
**Univ. of Texas at Arlington**

**CC-IN2P3, Lyon**  
**November 21, 2011**

# Outline



- Introduction
- The Current System
  - Direct Access in PanDA pilot (slides from Paul Nilsson)
- The Fun Part
  - Some ideas for fully Federated World
    - Mostly conceived this morning
  - How far can PanDA take us?

# Introduction



- Ongoing US ATLAS tests (Rob Gardner talk)
  - Learning pros and cons of various approaches
  - Going through many optimizations
  - Implementation in PanDA will depend on the results of this task force
- Using PanDA with FAX
  - Working already – but only for a single use case (direct access through local/global redirector)
  - More interesting – how to implement all workflows

# Direct Access in PanDA

- If site allows for remote file open, pilot will skip stage-in if the file is a root file and let the payload open it directly
- (1) Mechanism for user analysis jobs is activated by using the schedconfig DB
  - Old method, used for illustration as to what goes on behind the scenes:
    - `copysetup = setup_string^oldPrefix^newPrefix^useFileStager^directIn`  
where `setup_string` is some path to a setup file used by the copy tool, `oldPrefix` is the SURL prefix, `newPrefix` is the TURL prefix, `useFileStager` and `directIn` are booleans to control whether file stager and/or direct access mode should be used (file stager not relevant for this discussion, booleans are not internally exclusive so both are needed)
    - Example: `copysetup=setup.sh^gsiftp://osgserve04.slac.stanford.edu/xrootd/atlas^root://atl-xrdr.slac.stanford.edu:1094//atlas/xrootd^False^True`

'oldPrefix' is substituted with 'newPrefix' by the payload wrapper (runAthena/runGen) in the actual file path. The input file paths the pilot gets from the LFC, contains SURLs. The payload wrapper gives the TURL based file list to the actual payload that opens the files directly

# Direct Access in PanDA

- (2) Mechanism for user analysis jobs is activated by using the schedconfig DB
  - New alternative method:
    - pilot can create a TURL based PFC directly using lcg-geturls if copysetup is on the form: `copysetup = setup_string^useFileStager^directIn`  
(with `useFileStager = False`, `directIn = True`; there is no need to manually specifying 'oldPrefix' and 'newPrefix')
    - The payload wrapper will then use this PFC directly, and the payload will open the files directly as before
  - For non-root files, the pilot uses the normal copy tool (implemented in a "site mover" class) to stage-in the data
    - Decision making is done inside the site mover (abort stage-in if root file and direct access mode, continue if non-root file)
  - For production jobs, the direct access mechanism does not require any special copysetups like above. It is enough for the job definition to set `job.transferType = 'direct'`. The pilot will generate a proper copysetup (using the 'new' method above) as well as a TURL based PFC
  - The pilot has used direct access as described above for many years with user analysis jobs. Limited testing only with production jobs (less experience)

# Summary of Current System



- User analysis with XRootD Direct Access
  - Flexible system, lots of Direct Access experience
    - Ex. local redirector at SLACT2 & SWT2 many years
  - Site-by-site configuration of file access protocol
  - Now delegates TURL forming to lcg-geturls
- Production with Direct Access
  - Of limited use – no checking of site capability
  - Will fail if site is not setup correctly
  - Typically used for specialized (re)-processing

# Moving to Federated World



- From Direct Access to Remote Direct Access
  - Pilot use global redirector instead of local redirector?
  - But suffers from performance penalty if done blindly
  - Blindness not issue initially – jobs are sent to data
- Direct Access is not the only paradigm
  - Want to be able to use other access methods
  - Copy using Federated XRootD
    - Not a good idea to do this in the pilot – need to make PanDA server aware of Federated data

# Using 'FAX' in Brokerage (1)



- Consider production first
  - PanDA workflow sends small chunks of data (~20 files) to chosen site, waits for 'finished' callback
  - Part of data could already be at site (preferred)
- FAX could be used for data discovery
  - Currently PanDA brokerage uses DQ2 catalog
  - Brokerage could use global redirector instead
- FAX could be used for data copy
  - To complete the data chunk needed
  - Need callback when copying completed



# Using 'FAX' in Brokerage (2)



- User Analysis
- Get list of sites from Global Redirector
  - Similar to getting list of sites from DQ2
- Schedule jobs where data is local
  - No change to PanDA – same as current system
- What about PD2P?
  - Use copy-store or copy-cache

# More on PD2P



- What is PD2P
  - PanDA Dynamic Data Placement
  - Two copies of all files are distributed automatically in ATLAS – additional copies are made by PD2P
  - Extra copies at Tier 1 based on integrated usage
  - Extra copies at Tier 2 based on current demand
- FAX could be used directly by PD2P
  - We could use copy-store instead of DQ2 – but this is only swapping one system for another
  - Using copy-cache more interesting - reduces stress on DQ2 and asynchronous deletion mechanisms

# We need Callback



- PanDA relies on callback
  - To avoid polling
  - Always used for input data blocks for production
  - Not used for PD2P – but could be useful for immediate rebrokering
  - Can global redirector provide callback or use DQ2?

# Summary



- So far, no real changes to workflow
  - Production – move data as needed
  - User analysis – jobs go to data
  - Through DQ2 or direct interactions with FAX
- PD2P could benefit
  - If we use copy-cache
- Remote access
  - May need new brokerage implementation, if performance improves

# Some Questions



- Big question – DQ2 & FAX (Graeme's talk)
  - Should PanDA access FAX through DQ2?
  - If we access global redirector directly and everywhere, do we need LFC?
  - We will always need DQ2 for datasets
- Performance question
  - Under what conditions should we use remote direct access?
  - Will copy-cache help PD2P?