# Discussion on data transfer options for Tier 3

Tier 3(g,w) meeting at ANL ASC – May 19, 2009

Marco Mambelli – University of Chicago
marco@hep.uchicago.edu

# Resources (beside your own)

- OSG (Grid client and servers, e.g. SE)
  - Tier 3 liaison
  - Site operation
  - Software tools
- ATLAS (DQ2, Compatibility tests, Support)
  - WLCG-client
  - DQ2 Clients developers
  - DDM operation
- University?
  - hardware, cluster management, facilities
- Tier3 (Suggest solutins, Support)
  - ATLAS Tier3 Support

# Current data solutions for Tier 3

- DQ2 Clients
  - wide range of platforms supported
  - easy to install/setup (almost oneliner)
  - only as reliable as the underlying file transfer
  - slow: uses unprivileged paths (firewall, low bandwidth, low priority, …)

# Possible data solutions for Tier 3

- Improving DQ2 Clients
- Build a layer on top of DQ2 Clients
- Install DQ2 Site Services or part of it

# Improving DQ2 Clients

- use FTS for file transfer
  - still data sink (invisible to the Grid, no consistency problem)
  - allow queuing
  - can use privileged paths
  - may require FTS channel setups
  - requires SE (server installed at the Tier 3)
  - development by DQ2 Client developers
- Storage Element
  - requires a server
  - tested path for data transfer
  - data will be copied there
  - provided by OSG (easy install, possibly VM?)

# Build a layer on top of DQ2 Clients

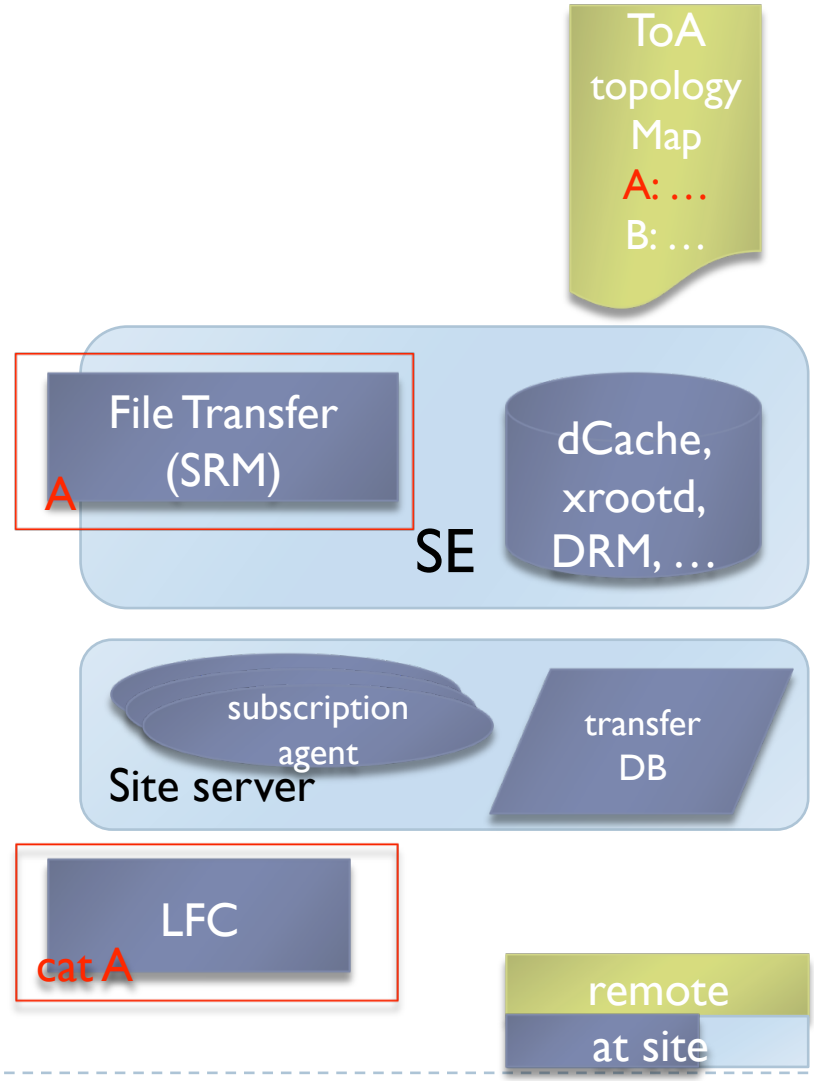
- manage retries
  - multiple attempts
  - try different 'copy tools'
  - still data sink (invisible to the Grid, no consistency problem)
  - no server required
  - unprivileged path
  - no FTS setup, not distinguishable from other clients

# Install DQ2 Site Services or part of it

- Full Site Service installation (Tier 2, Alden's VM)
- Hosted Site Service or DQ2 End Point (Non US ATLAS Clouds)
- Something inbetween
  - installation can be made simple
  - consistency problem
- 'Data Sink' Site Service
  - remove consistency requirement
  - DQ2 development

# Separate Elements of DQ2 (Site) Server

- (TiersOfATLAS ToA)
- Storage Element
  - [managed] disk space
  - transfer server
- DQ2 Endpoint
  - entry in ToA
  - FTS channels
- DQ2 Site Services
  - transfer agents
  - transfer DB
- LHC File Catalog (LFC)

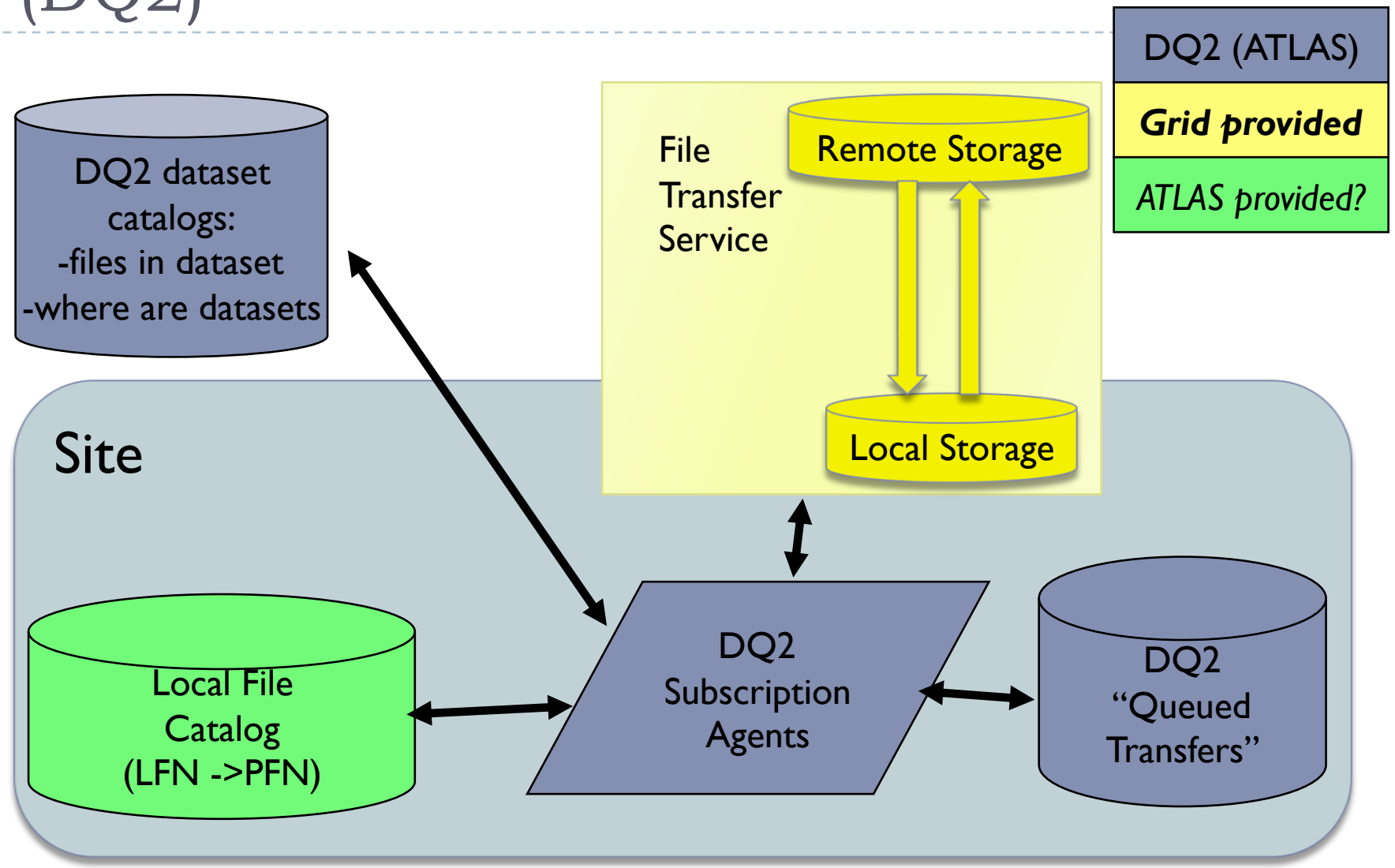


# Storage Element

- Disk pools
- Storage system
  - DRM, dCache, GPFS, Xrootd
  - uniform space (file retrieval and naming)
  - file system abstraction
- External access and file transfer
  - GridFTP (GSI, multiprotocol)
  - RFT, FTS (reliable)
- SRM Storage Resource Management
  - space management
  - transfer negotiation and queuing
  - pinning
- File catalog (RLS, LFC, RLI, LFC)

# END – Extra follows

# ATLAS Distributed Data Management (DQ2)



# Distributed Data Management (DDM) capabilities

- Different setups/components presented
- Incremental setups (each one requires the previous)
- Resulting configurations (adding the components)
  - c0ddm: no local managed storage, relying on external SE
  - c1ddm: SE only (your ATLAS visible files are somewhere else)
  - c2ddm: DQ2 endpoint and SE (Site services and LFC outsourced) (UofC Tier3)
  - c3ddm: DQ2 site services + endpoint + SE (LFC outsourced)
  - c4ddm: LFC + DQ2 ss + endpoint + SE (US Tier2)
- Each time the difficulty to add and maintain the additional component is evaluated
- Requirements, functionalities and responsibilities are listed

# No local Storage Element

- Copy to the local disks
  - any local path (also fragmented, user space, …)
- DQ2Clients (enduser tools)
  - dq2-get/dq2-put
  - limited automatic retry
  - no automatic replication (conditions)
  - synchronous (no queuing)
  - immediate
  - unprivileged data paths
  - chaotic
  - strong support (VDT, ATLAS)
  - only intermittent outbound connectivity

DIFFICULTY RATING – Easy to install (user application), many supported platform, simple use, documented, strong VDT and ATLAS support

# No local Storage Element - **FTS**

- Copy to the local disks
  - any local path (also fragmented, user space, …)
- DQ2Clients (enduser tools – **development required**)
  - dq2-get/dq2-put
  - limited automatic retry
  - no automatic replication (conditions)
  - synchronous (no queuing)
  - immediate
  - **privileged data paths**
  - **organized**
  - strong support (VDT, ATLAS)
  - **requires OSG SE,** only intermittent outbound connectivity

DIFFICULTY RATING – Easy to install (user application), many supported platform, simple use, documented, strong VDT and ATLAS support

# OSG Storage Element

- Big variety of Storage Elements (SRM/unmanaged)
  - Gridftp server
  - dCache installation with SRM and Space Token
  - (dCache, BeStMan-Xrootd, BeStMan-FS, Gridftp)
- Help from OSG (and US-ATLAS)
  - software package
  - support (tickets and mailing lists)
- You have to do the installation and maintain it functional
  - specially if you want a DQ2 End Point or more
- You have a **standard file server**

Could be very easy, depending on the FS and SE chosen. Some file server are difficult to tune/maintain. You may have them already. OSG/ATLAS do not add load (compared to local use only)

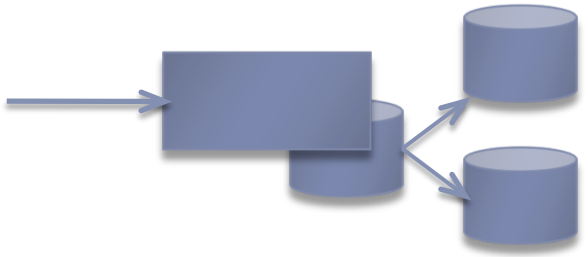
# DQ2 End Point

- Your storage element is visible in DQ2
- Can use **DQ2 subscriptions** to transfer data
- You need the support of a Site (Tier2) with
  - DQ2 Site Services
  - LFC
- Register your SE in TiersOfATLAS
  - Setup and maintain FTS channel (support and effort from BNL)
- Functional and available SE
  - (May have to support SRM or StorageTokens)
- Reliable data management
  - Your files are known on the Grid through DQ2
  - Designate a data manager

Excluding the SE, more bureaucratic than technical. Outside support would help

# DQ2 "sink" End Point (Subscription only DQ2)

- A storage element able to receive subscriptions (in ToA)
  - one or 2 level staging
- The SE is not visible in DQ2 as source of data
  - no LFC or registrations in central catalog
- Can use **DQ2 subscriptions** to transfer data
- No data management responsibilities
- **Does not exist**
  - unknown development effort
- Different scale in the number of endpoints (load?)
  - unsustainable strain on Tier1/2 (SE, network)
  - complicate network topology (FTS paths)
  - huge number of subscription



**?** Unknown development effort. Strain on the system (Tier1/2, network)

# DQ2 Site Services

- You move your own data
  - DQ2 agents are running at your Site
  - Do not rely on other Site being up (for file transfer
  - Install and configure Site Services
  - Maintain reliable Site Services
- You need a Storage Element and a DQ2 End Point
  - Still need FTS (support and effort from T1)
- You need a LFC
  - From a Tier2/Tier1 or your own

Optimistic? Could require effort: software tested more on LCG, some problems in the past

# LHC File Catalog

- ATLAS migrating from LRC to LFC
  - complex transition problem
- You have your local catalog
  - No need of a network connection to know which files are at your Site
- Generally it makes sense if you have also a SE, DQ2 End Point and DQ2 Site Services
  - Independent from another Site (Tier2)
  - Still need FTS (support and effort from T1)
- You have to install, configure and maintain LFC
  - Database back-end, LFC Server program

Difficult. LFC, is relatively new, some load problem, maintain consistency

# More classification examples

- US ATLAS (OSG) Tier2, WISC (Tier3)
  - c4ddm, c2je
- LCG Tier2
  - c3ddm (LFC hosted at supporting Tier1), c2je (gLite)
- UofC Tier3
  - c2ddm, no CE (local PBS queue)
  - shares some disk (OSG_APP, OSG_GRID) with the Tier2
- UIUC, Satellite clusters (UC_ATLAS_Tier2, IU_OSG, OU_ATLAS_SWT2)
  - c0ddm, c2je

# Documentation

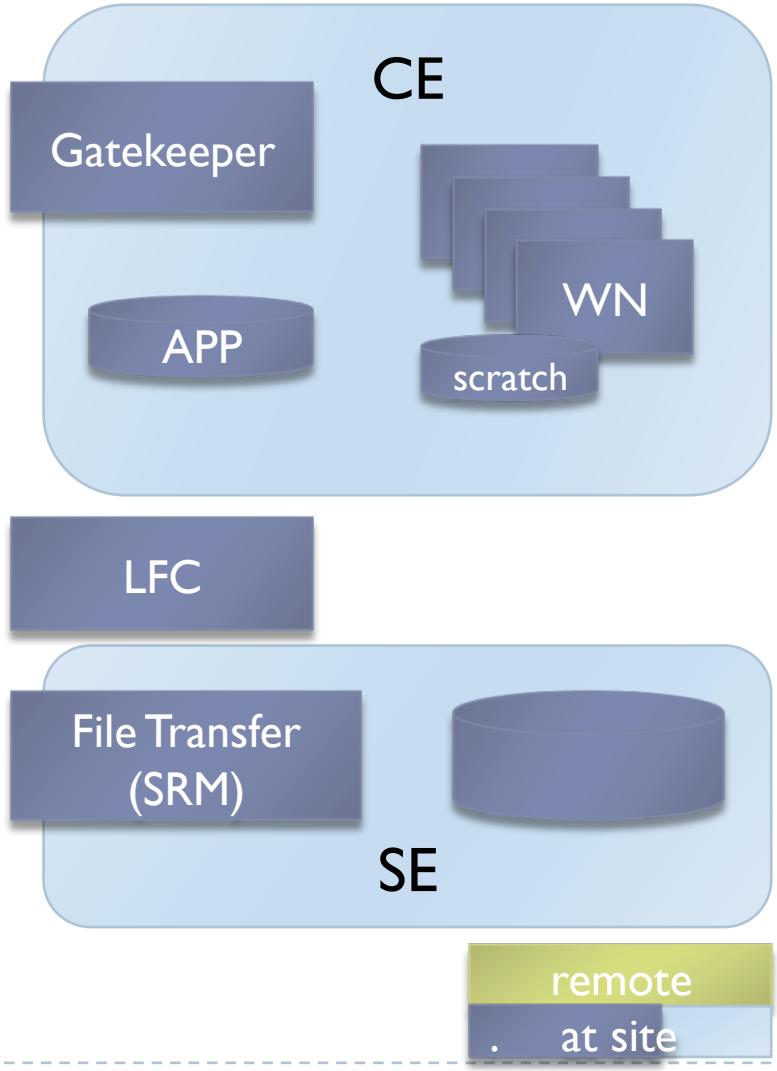
- DDM
  - https://twiki.cern.ch/twiki/bin/view/Atlas/DistributedDataManagement
  - https://twiki.cern.ch/twiki/bin/view/Atlas/PandaDataService
- TiersOfATLAS
  - https://twiki.cern.ch/twiki/bin/view/Atlas/DDMTiersOfATLAS
- LFC
  - https://twiki.cern.ch/twiki/bin/view/LCG/LfcAdminGuide
- Clients
  - https://www.usatlas.bnl.gov/twiki/bin/view/Admins/WlcgClient
  - https://twiki.cern.ch/twiki/bin/view/Atlas/DQ2Clients
  - https://twiki.cern.ch/twiki/bin/view/Atlas/DQ2ClientsHowTo

# A Tier3 example

- Install OSG Storage Element on top of the data storage and setup a DQ2 End Point (c3ddm)
  - Big datasets are easily moved with subscriptions
  - Little effort for the Tier3
  - Supporting Tier2 and Tier1 can handle most of the effort (and they should handle it easily once the procedure is streamlined)
- Install if desired an OSG Computing Element (c2je)
  - Specially if it has more than 10 computers (cores)
  - Effort still little
  - Can receive ATLAS releases
  - Can use its resources with Panda/pAthena (and can contribute its idle CPU)

# CE elements interacting with Panda

- Gatekeeper
  - receive pilots (authenticate, receive files, schedule)
- Worker Nodes
  - execute pilots and jobs
  - ex. installation jobs
- Shared file systems
  - host ATLAS releases (APP)
  - support execution
- (SE and LFC)
  - receive/provide files from/to Panda Mover



# Job execution capabilities

- To interact with ATLAS is recommended a workstation where you can run the ATLAS sw (Athena, ROOT, …), Grid client packages (WLCG-Client, gLiteUI, …) and DDM client (DQ2Clients)
- To execute jour own (or someone else's) pAthena jobs you need a system capable to run Panda pilots
  - c1je: Grid CE
  - c2je: Grid CE + additional Panda support (Panda site)
- In the following pages I'm referring to an OSG CE. LCG/ gLite has a similar solution to provide for a CE

# OSG Computing Element

- Help from OSG (and US-ATLAS)
  - software package
  - support (tickets and mailing lists)
- You have to do the installation (and maintain it functional)
- If you don't give your availability for Panda you have no obligation or responsibility
- You can receive ATLAS releases (Grid installation)
- You are part of OSG
  - Can use the grid to reach your cluster, run on it
  - Can be in Panda/Pathena and in ATLAS accounting

Software is well documented with a large user base, almost easy

# Panda support

- You can submit pAthena jobs to you own resources
- Until GlExec is deployed you cannot discriminate between Panda users
  - limit external use by not advertising
- Satisfy other ATLAS jobs requirements
  - access external HTTP repositories
  - access external DBs
  - run ATLAS sw
- You need an OSG Computing Element and a c4ddm (or the support of one
  - old Panda pilot (no forwarded proxy) could run also on local queues but this requires extra support

Depending on your configuration it may be difficult to support ATLAS requirements (specially if you are behind a firewall)

# ATLAS space token

| token name | storage type | used for | @T2 | @T1 | @T0 |
|---|---|---|---|---|---|
| ATLASDATATAPE | T1D0 | RAW data,<br>ESD, AOD from re-proc | | X | X |
| ATLASDATADISK | T0D1 | ESD, AOD from data | X | X | X |
| ATLASMCTAPE | T1D0 | HITS from G4,<br>AOD from ATLFAST | | X | |
| ATLASMCDISK | T0D1 | AOD from MC | X | X | X |
| ATLASPRODDISK | T0D1 | buffer for in-and export | X | | |
| ATLASGROUPDISK | T0D1 | DPD | X | X | X |
| ATLASUSERDISK | T0D1 | User Data | X | X *) | |
| ATLASLOCALGROUP DISK | T0D1 | Local User Data @T3 | | | |

How data is organized at sites (SEs)