Creating Federated Data Stores For The LHC

Summary

Andrew Hanushevsky
SLAC National Accelerator Laboratory

November 21-22, 2011
IN2P3, Lyon, France

Definition: Space Federation

- We need a common definition
- Collection of disparate storage resources managed by co-operating but independent administrative domains transparently accessible via a common name space.

Data Access Models

- Transparent replication usage
- Opportunistic access (additional access modes)
  - CMS is leading on this
- Support may become more complicated
- Latency issues come up and can be problematic
- Proxy access adds overhead needs improvement

Data Access Models

- Everyone is interested in direct access
  - Perhaps direct access more suited from bigger sites
    - Disagreement on role of smaller sites
  - Access must be optimized based on client/server location
  - Alice already doing this in production
  - ATLAS and CMS are actively exploring this

Data Access Models

- Caching
  - Should be more akin to catalog realignment (healing)
    - I.E. Making storage contents consistent with the catalog
    - Alice uses GLRD to provide data location consistent with catalog
  - Opportunistic caching
    - There is interest but may be very problematic
      - Alice does not use opportunistic storage
      - Atlas would like a call-back to track such placemen
    - This may be more relevant for Tier 3 sites
      - But no one wants to micro-manage the site
      - Though the contents should be discoverable and usable
  - Monitoring and more experimentation essential

Data Access Models

- Federated "bad" servers can impinge on everyone
  - Need alarms and active request monitoring
  - Clients should be able to have bandwidth minimums
    - Automatically switch to a better source or
    - Use multi-source access for block reads
      - Efficiency issues exist
  - Generally, at any point something is going wrong
    - Need better avoidance of "wrong" places
    - Need better dynamic selection of "right" places

Data Access Models

- Clustering/Federating Cloud Storage
  - Clouds needs a general storage access solution
  - Private vs public cloud probably a mixture of both
  - Cost structure is key
    - Could be more than 10x more expensive w/ I/O-Storage charges
  - This is bleeding edge and work just started

Name Spaces

- CMS Static name mapping (lfn to pfn)
- Only Atlas uses "LFC"
  - Others have a simple lfn to pfn mapping or use dual global names
- Atlas has no deterministic global name space
  - Yes, LFNs are "global" but not actual access path
  - Problems will be addressed in Rucio much like CMS
- Alice externalizes LFN but remote access via PFN
  - A Dual global name space has certain advantages
    - Eases catalog realignment and space reclamation
- Bottom line – Federation efficiency requires a deterministic global name space.
- The best approach is to keep it simple and stupid.
  - There should be no expectation that the space is browsable

Federation Mechanisms

- Other protocols can also be used to federate
  - E.G. http has some built-in end-user appeal
  - But will likely require significant augmentation
    - To provide robust and flexible federation
  - Even then may or may not be sufficient
  - Additional studies are needed
- dCache rapidly moving to ease federation support
  - Plug-in architecture, easier deployment

Monitoring

- Monitoring data into root or Hbase for analysis
  - Move to actively augment traditional RDBMS
- Considered for use in file popularity selection
  - Also in file caching and purging decisions
  - Also to identify bad files
- Access to real repeatable tests are still lacking
- What is the ideal standard metric?
  - Is the key metric event rate?
- We are still discussing many things
  - Record formats & aggregation levels
  - Common infrastructure

Goals

- Identify impediments to successful federation
  - Must avoid bad apples in the federation
  - Site selection needs to be more clever at global level
  - More monitoring information (especially client-side info)
  - Inclusion of more sites not running "boxed" xrootd
  - Understanding the actual use modes
    - E.g. EOS, better integration of dCache (evolving), etc
  - But successes should not be ignored
    - There are actual working federations!
      - Alice in production
      - CMS and Atlas nearing production

Goals

- Outline broad technical solutions
  - Better site selection at global level
  - Perhaps use out-board information for site selection
    - Optimization handled by some external oracle?
      - DDM, Alice catalog, etc.
  - Global Redirector can be used as a fallback when all else fails
  - Monitoring is the key to making this effective
  - Needs to be better explored how to best implement this
    - This is a common problem regardless of federating mechanism

Goals

- Outline broad technical solutions
  - Identified broad monitoring information areas
    - Started fleshing out additional metrics (Servers & Clients)
      - Need to solidify the list
    - Actual aggregation is still an open issue
    - Report record formats need to be standardized
      - Use some existing but usable format (e.g. WLCG? , OGF?)
  - Overlap still exists
    - E.G. Collector, packet parsing
  - No clear path to reduce redundant effort
    - Perhaps an *active* monitoring work group
      - In any case, a monitoring package will need to be maintained

Goals

- Outline broad technical solutions
  - Bandwidth driven client source routing
    - Actual waiting for new client
  - Deterministic namespaces
    - Alice and CMS already have them
    - Atlas on the road and Rucio may be the answer
  - Plug-in architecture for dCache
    - Ongoing

Goals

- Establish framework for technical co-operation
  - We have started doing this by virtue of this meeting!
    - Should it be protocol focused or broad-based in the future?
  - Do we need a more structured approach?
    - Perhaps a list specifically for this topic (federations-l)?
  - Overlap in monitoring activities
    - Should there be a separate cooperative framework for this?
  - Reduce redundant effort
    - This will happen naturally if we have a framework

Goals

- Ideally, spur the adoption of federated storage
  - Conservatively, it looks like it is jelling
- So, should we have another meeting?
  - Yes, once again at IN2P3 within a year

Thank You

- IN2P3
  - For hosting this meeting
- Jean-Yves Nief
  - For local organization & meeting web page
- Stephane Duray & administration team
  - For excellent logistical support