Memorandum

ATLAS Datasets Subscription Policy

To :   M.Branco,P.Salgado, DQ2 SW developers
From : S.Campana,K.De, A.Klimentov
CC :   D.Barberis, K.Bos, DDM Operations Group

September 28, 2007

.

In this memo we want to address the subscription policy for Production and ATLAS data . It is based on recent observations during DDM operations and inefficiency of the existing policy. The subscription policy and replication strategy of User's datasets is not covered by this document. We also assume that at least one Tier-1 has dataset replicas.
The policy may be optimized overtime, and the document will be revised after consultation with all communities.

General remarks

.

It is proposed do not support Tier2(s)-Tier2 data transfer for tiers belong to different clouds. Tier-2 centers must subscribe data from the parent Tier-1. Individual requests for bulk data transfer should be placed using DDM data transfer Web interface or DDM Operations Savannah page. Individual Tier-2-Tier-2 subscriptions for centers in different clouds will be blocked by the system.

1. DQ2 subscription processing

1.1. Dataset subscribed without specifying source location.

- Destination site is a Tier-1 center :
    - check Tier-2s in the same cloud and subscribe from Tier-2s if replicas is available
    - subscribe from all Tier-1s with the complete replicas
    - subscribe from all Tier-1s with incomplete replicas

– subscribe from Tier-2s with replicas

The last item (subscription from Tier-2s from the foreign clouds) is considered as failover in case data cannot be obtained from Tier-1s (or Tier-1 is unavailable) during 1 week (*TBD*).

- Destination site is a Tier-2 (Tier-3) center :

  – subscribe from parent Tier-1

  – subscribe from Tier-2s in the same cloud

  – subscribe from all Tier-1s with the complete replicas

  – subscribe from all Tier-1s with incomplete replicas

  – never subscribe from Tier-2s in different clouds

Items 3 and 4 assume non-hierarchical data transfer, and not completely agree with ATLAS Computing Model, the implementation requires that Tier-1s allow data to be pulled by any Tier-2 in ATLAS. Tier-1 has right to cancel subscriptions in case of heavy loading or/and Storage Element (SE) performance degrading to prevent problems for its cloud.

1.2. Dataset subscribed using *source* **option.**
It is mandatory to subscribe dataset only from the sites listed as *'sources'*. The subscription priority should be ordered according to the list below :

- the nearest Tier-1 (if it has a complete replicas)

- all Tier-1s owners of the complete replicas

- all Tier-2s from the same cloud

- all Tier-1s owners of incomplete replicas

if destination site is Tier-1, then it can be subscribed from foreign Tier-2s listed in *'source'* option, only in the following cases :

- none of Tier-1s has a complete replicas

- file(s) transfer from Tier-1(s) to Tier-1 is failed after several attempts (the number of attempts TBD). If destination site is Tier-2 it can be subscribed only from Tier-2s in the same cloud.

2. Automatic datasets replication

Application programs for automatic data distribution (AOD, validation samples, conditions data and User's requests) will follow the logic below :

- AODs are subscribed from Tier-1 to Tier-1s and CERN.

- Validation samples are subscribed from all available sources to the pre-selected Tier-1s

- If user's request is approved by regional DDM coordinator, then datasets are subscibed follow to the algo

    - Tier-2 is subscribed from the parent Tier-1
    - if data isn't available on Tier-1 then Tier-1 is subscribed from all avaialble sources and DQ2 will follow the algorithm described in subsection 1.2 of current document.

3. Cosmics datasets replication

Subscription policy and data transfer must follow ATLAS Computing Model.