Data selection from metadata catalogues

DataGrid WP9

Final DataGrid Conference

September 26th – October 1st 2003

Heidelberg

Data selection from metadata catalogues

- With the previous version Data Management System (DMS) we needed to:
  - query our Metadata catalogue outside the Grid (using spitfire) to get the LFN needed
  - submit the job on the grid.
  - If needed, insert metadata in a catalogue outside the Grid (using spitfire).
- With the new Replica Metadata Catalogue (RMC) inside the RMS we can elaborate the 3 following propositions :
  - Catalogues on the Grid (Will it be possible?)
  - Using the RMC Attributes
  - Using the RMC Aliases

Catalogues on the Grid (Will it be possible?)

- Is it possible to add to the RMC the existing metadata catalogues, developed in WP9?
- If so, a job, submitted on the grid, will query the concerned WP9 metadata managed by the RMS to extract a list of SFNs that matches the requirements.
- Advantages :
  - To have an easy data selection on the Grid
  - To avoid additional selection steps outside the Grid
- Drawbacks : none

Using the RMC Attributes

- Addition to the RMC of the WP9 relevant attributes
- Advantages :
  - To have an easy data selection on the Grid : a job submitted will query the RMS by means of relevant attributes to extract a list of GUIDs
  - To avoid additional selection steps outside the Grid
- Drawbacks :
  - All the attributes are available for all the data of our VO. (Attributes needed for Level1-GOME data will be left blank for Lidar data: using this method, 18 different attributes will be added to the VO metadata catalogue)
  - As for now attributes can only be strings, parsing routines will have to be written to allow for queries on date and geolocation.

Using the RMC Aliases

- Elaboration of conventions on aliases.

  For one of our application the key parameters are the date and the geolocation. One of the possible conventions could be:

  sensor_site(or location: latitude-min_latitude-max_longitude-min_longitude-max)_parameter_date

  (LIDAR_OHP_O3_1999_08)

- Advantages :
  - To make selection only by means of the filename (quick selection)
  - To avoid additional selection steps outside the Grid
- Drawbacks :
  - Limited Length of the alias's LFN
  - Need to elaborate convention

What WP9 really need

- Integration of Spitfire functionality into RMS.
  - to create, modify, delete or drop tables in RMS (regarding authorisation), containing the WP9 metadata attributes
  - to avoid metadata storage outside the Grid.
- Spitfire still needed for accessing other databases than metadata catalogues.
- Both Spitfire and RMS with the same API or CLI

  Example:
  - *edg-rms-query-for-guid –**source=RMS** –catalogue=GOME_OPERA –query "lat between 0 and 50 and lon between 0 and 90 and date between 2001-01-01 and 2001-01-02"*
  - *edg-rms-query-for-guid –**source=datagrid.nadc.nl** –catalogue=GOME_OPERA –query "lat between 0 and 50 and lon between 0 and 90 and date between 2001-01-01 and 2001-01-02"*

Conclusion

- To be discussed with WP2
  - We have discussed 3 possible solutions: Which one is the most feasible according WP2?
  - Is it possible to integrate our Global database or our distributed databases into the RMC?
    - If yes, is there any limitation?
    - Is the RMS able to do the interface?
    - Is it possible to set up restricted access?
    - Is it possible to duplicate this database located in RMC?
  - What is the maximum length of a LFN?
  - Are the RMC attributes valid
    - for all the files in the RMS belonging to a given VO
    - or is it possible to make collections of files and give them special attributes?
  - What we really need is Spitfire functionality integrated in RMS. Will this be the future?