Grid Tools

Replica/File Catalog based on MAGDA @ RAM Data @ SUNY

Shevel@ram0.i2net.sunysb.edu

10 Dec 2002

Our needs in Replica/File Catalog

Ü **Our Conditions**

  Ü Relatively small physics team (about 10 persons).

  Ü Needs to keep some book-keeping to keep information about our files (mainly flat files) in different locations (in our computers, at BNL, etc.). Expected total number of files is about $10^{**}5 - 10^{**}6$ (now is about $2*10^{**}4$)

  Ü Needs to keep the catalog more or less up to date.

Our needs in Replica/File Catalog (continuation)

Ü **Our Conditions**

- Ü From time to time we need to transfer a group of files (from about 10**2 to 10**4 files) in between different locations (in between SUNY and BNL).

- Ü Apparently we need to keep newly copied files in Replica/File Catalog. Some trace of our data transfers is required as well.

- Ü Finally we need to have program system or several stable systems for all of mentioned tasks.

MAGDA

Ü Several replication systems are available now for High Energy Physics. We are testing MAGDA.

Ü MAnager for Grid-based DAta = MAGDA (http://www.atlasgrid.bnl.gov/magdadoc/userguide.htm).

  Ü MAGDA consists of many scripts in perl, kernel of the catalog is based on MySQL.

General terms

Ü REPLICA - a COPY of a file in DIFFERENT PLACE to MAIN LOCATION.

Ü MAIN REPLICA (Master COPY, MAIN COPY) – file in MAIN LOCATION.

Ü A DIFFERENT PLACE   - might mean different directory, different media or different host.

Ü VIRTUAL ORGANIZATION (VO) - name of any group: one person, small team, detector group, University team, collaboration.

General terms (continuation)

Ü FULL LOGICAL FILE NAME - means line in format:
lfn:/VO/short_logical_file_name

Ü SHORT_LOGICAL_FILE_NAME   – name. This name is the same in spite of location.

Ü PHYSICAL FILE NAME   - means line in format   pfn:/HOST/PATH/file-name

General terms (continuation)

CONSISTENCY - means that File/Replica Catalog does contain up to date information.

shevel@ram0.i2net.sunysb.edu

MAGDA terms

Ü **COLLECTION – group of files.**

Ü **TASK – operation of replication of the COLLECTION.**

Ü **SITE – the place where data are.**

Ü **LOCATION – precise directory where data are.**

Ü **HOST – host where MAGDA clients might be run**

RAM Prototype for File/Replica Catalog

Ü **Now we have to go to see real prototype**

Ü **The prototype is located at**
   http://ram3.chem.sunysb.edu/magda/dyShowMain.pl

CONCLUSION

Ü **We are playing now with the prototype of the File/Replica Catalog.**

Ü **First opinion it is very useful tool understand better what do we need in reality.**

Ü **We discovered that such catalog is very helpful for any kind of files (data, programs, papers, pictures, etc.).**

shevel@ram0.i2net.sunysb.edu