



Enabling Grids for E-science

SAM Archiving Requirements

www.eu-egee.org



- **SAM framework stores hourly test results for all production Grid services and from different VOs.**
- **For each test result we store (TESTDATA table):**
 - VOID (F.K. to VO table: under which VO the test run)
 - TESTID (F.K. to TESTDEF table)
 - NODEID (F.K. to NODE table: which host was tested)
 - ENVID (F.K. to TESTENV table: group tests under 1 submission)
 - STATUSID (F.K. to TESTSTATUS table: OK, ERROR, etc)
 - MAINTID (F.K. to MAINTENANCE table: declared downtimes)
 - TIMESTAMP (when the test run)
 - INSERT_TIME (when the test entered the DB)
 - SUMMARYDATA (255 VARCHAR2: summary of output)
 - DETAILEDATA (CLOB field with test STDOUT/STDERROR)
- **This means around 40 GB of new data per month.**

- **The CLOB is highly-compressible ASCII output, but it's the culprit in terms of DB space**
- **For performance reasons, it's desirable to remove old CLOBs – but context is needed, so lots of other tables need to be saved too**
- **125GB Oracle export file created, which we would like to store off-line (Castor or TSM, we don't mind).**
- **Proper retention policy is needed:**
 - Old data will probably never be accessed, but if we delete it, someone will want it 😊
 - MB approval will be needed – especially if we decide to delete

SAM test: *CE-sft-csh*
 Submitter VO: *ops*
 Node: *lcgce01.cpp.ualberta.ca*
 Execution time: *10-Nov-2008 15:11:55*

Checking if CSH works

CSH script:

```
#!/bin/csh
env | sort > env-csh.txt
```

Testing if CSH script works:

```
+ ./csh-test.csh
+ set +x
```

CSH script generated the following output:

```
APEL_HOME=/hepusers/atlas/atlas.glite/glite-WN-3_1_4-1/glite
ATLAS_CONDDB=ca
_=./csh-test.csh
DCACHE_CLIENT_ACTIVE=1
DCACHE_REPLY=thuner-gw.cpp.ualberta.ca
DQ2_GSIFTP_HOST=gsiftp://lcgse00.cpp.ualberta.ca:2811
DQ2_LFC_HOME=/grid/atlas
DQ2_LOCAL_ID=ALBERTA
DQ2_LOCAL_PROTOCOL=dcap
DQ2_SRM_HOST=srm://lcgse00.cpp.ualberta.ca:8443
DQ2_STORAGE_ROOT=/pnfs
DQ2_URL_SERVER=http://atldmcat.cern.ch/dq2/
DQ2_URL_SERVER_SSL=https://atldmcat.cern.ch:443/dq2/
DQ2_USE_SRM=False
DYLD_LIBRARY_PATH=/hepusers/atlas/atlas.glite/glite-WN-3_1_4-1/globus/lib
EDG_LOCATION=/hepusers/atlas/atlas.glite/glite-WN-3_1_4-1/edg
EDG_WL_JOBID=https://wms201.cern.ch:9000/K2v1KYbU7U0tXRBokct02w
EDG_WL_LOCATION=/hepusers/atlas/atlas.glite/glite-WN-3_1_4-1/edg
EDG_WL_SCRATCH=/localdisk/lcg
ENVIRONMENT=BATCH
```

- **Would like to archive all test results that are older than 24 months. This means all of these tables:**

<p>TESTDATA VO TESTDEF NODE TESTENV TESTSTATUS MAINTENANCE</p>	<p>SERVICE SITES ORGANISATIONS</p>
----------------------------------------------------------------------------------------	--------------------------------------------

- **Also, we don't need the CLOB column (test STDOUT/STDERROR) of TESTDATA table, after 6 months.**
- **MB will decide whether archive means delete or save to off-line storage**

- **Possibly redesign TESTDATA table to contain pointers to a CLOB table (deletion & reclaiming space is easier)**
- **Currently, without history of Grid Topology, the value of old test results is debatable (so why keep them?)**
- **New topology database with history is being designed, so archiving will be useful in one form or another**
- **We feel that this is a service that the DB group should provide!**