Table of Contents

Issue: Resource Broker Disk Space requirements

Problem Statement

The resource broker stores sandboxes for the user's jobs (input and output).

A review of LHC capacity requirements is as follows

Using the data presented during the meeting on Monday, we arrived at

- 10 MB input sandbox
- 10 MB output sandbox
- 14 days retention time of sandbox data

The current LSF batch throughput at CERN is 906 jobs/hour. Therefore, we would have 304416 jobs during a 14 day period on the current load if all jobs were submitted through the grid. We are expecting a 4-5 times growth of CPU capacity for LHC which would lead to at least 3 times more jobs, i.e. around 900,000 sandboxes.

With 20MB per sandbox, this is 18TB of disk space for all RBs.

While this seems high, I cannot find any errors in my calculation.

Maarten's views

Your figure corresponds to a worst-case scenario, in which no jobs are cleaned up during two weeks. I would expect a job normally to be cleaned up within a day after it has finished, in which case 2 TB would suffice. However, RBs will also submit to other sites, so the amount of space for CERN jobs should be multiplied by some factor. On the other hand, it cannot be just the RBs at CERN that drive the grid: there will be RBs at other (big) sites too, probably scaling with their computing resources. All in all, some fudge factor is needed, but it need not be as high as 10. I think that 18 TB would be fairly conservative; half of that may suffice.

Distribution to LCG Rollout

Publication from : Maarten Litmaath 1689 <Maarten.Litmaath@cern.ch> (CERN) This mail has been sent using the broadcasting tool available at http://cic.in2p3.fr⊠

Dear colleagues, in the past 2 weeks there have been serious problems with CERN production RBs due to file systems filling up completely with huge output sandboxes. The worst example:

total 59492104 -rw-rw---- 1 cms002 edguser 60860395520 Sep 18 03:36 ORCA_000097.stderr -rw-rw---- 1 cms002 edguser 13778 Sep 17 16:32 ORCA_000097.stdout

Indeed: a 60 *GB* file! Filled with the same error message over and over.

Obviously we need to do something about it fast.

The next version of the RB code, currently being tested, will limit the size of an output sandbox to a maximum value set by the RB admin.

The job wrapper sorts the files in the output sandbox by size and copies those files to the RB whose combined

size does not exceed the limit; the difference between the combined size and the limit is divided by the number of remaining files and each such file is truncated to the resulting value, after which it is copied to the RB. An event is logged for each file that needed to be truncated or was not found. In that case edg-job-status will show that the job is \"Done (with errors)\" and as usual edg-job-get-logging-info -v 1 will have the details.

Each output sandbox globus-url-copy to the RB is tried in a loop: if it fails, the problem is assumed to be temporary (e.g. network down) and the operation is retried after a delay that is doubled each time, starting at 5 minutes; the job wrapper will give up after 5 hours. An event is logged for any globus-url-copy problem.

The maximum output sandbox size should be set to a small value, e.g. 10 MB like for the input sandbox. An RB is *not* an SE. However, to smoothen the transition we propose to start with 100 MB.

To mitigate the problem on the RBs right now we have launched a continuous cleanup job with the following characteristics:

- any sandbox file older than 3 weeks is deleted;

- any sandbox file larger than 100 MB is truncated to 100 MB;

- any sandbox file larger than 10 MB whose name matches the following patterns is truncated to 10 MB:

*.out *.err *.log *.stdout *.stderr

Comments?

Date	Who	Description
16/09/2005	Tim	Mail exchange with Maarten

-- TimBell - 16 Sep 2005

This topic: LCG > IssueRbDiskSpace
Topic revision: r2 - 2005-09-21 - TimBell