egee

Enabling Grids for E-science in Europe

www.eu-egee.org

IT/CZ cluster meeting, Milano, September 23–24, 2004

Addressing a few Known WMS Issues with L&B-based Hacks

Aleš Křenek et al.

- Several serious issues on WMS reliability raised recently
  - Usually tracked to problems outside "pure" WMS
  - Perceived job failure percentage is the primary criterion
  - Damage the WMS perception among users
- Fast reaction required

Suggested Approach

- Not enough time to provide clean, well-designed solutions

- Pragmatic "hacks" possible—not completely reliable but good chance to improve statistical behaviour

    - And end users' perception of the overall behaviour

- Most of run-time information required to improve WMS behaviour is available in L&B for zero or little cost

Failures on stage-in

- Download performed by JobWrapper fails *before* starting the job payload

- Resubmission is disabled—the user really does not want the payload job to be restarted

- Current result: an overall failure of the job

- Solution:

  - Log another *Done-FailedDownload* L&B event from JobWrapper when only a download fails

  - If this event is seen by LM, resubmit the job, regardless of the setting of the true retry count

Incomplete/out-of-date CE information

- An advanced LRMS configuration imposes a restriction (e.g. the number of running jobs of a VO) but this is not reflected in the information system (e.g. due to a schema)

- Information systems do not reflect changes fast enough in the case of bulk submissions

- Current result: suboptimal job distribution or even unnecessary failures

- Solution:

  - (Mis)use JobWrapper of an already running job as a spy gathering up-to-date information on the CE, log as L&B events
  - Unreliable, empiric method to retrieve the info (e.g. try PBS `qstat` and do not care if it is not found)
  - Process the events on L&B server to give per-CE values (slide 6)
  - Use them in WM to adjust CE ranking (at least)

Black holes

- Misconfigured site accepts jobs at a high rate, most or all of them fail
- Even false success may be reported
- Solution:
  - Compute the time between starting a job and its failure (and successful completion) in L&B, include among the per-CE L&B data
  - Encourage users to specify estimated job execution time, gather it as a L&B user tag, and compare with reality
  - Adjust CE ranking accordingly, or generate a (temporary) CE blacklist

Processing per-CE data in L&B

- After processing an incoming event, L&B server calls one or more hooks

- The hook function extracts or computes an appropriate quantitative information (e.g. time before job failure) and stores it (organised per-CE) in a *Round-Robin Database* `http://people.ee.ethz.ch/~oetiker/webtools/rrdtool/`

- Data from RRD are made available to WM

  - directly, as shared files
  - over network protocol (LDAP, additional L&B query, . . .)

- Some of the serious issues raised by the WMS users recently can be addressed with pragmatic "hacks" quite easily

- No clean and completely reliable solution, but it can improve statistical WMS behaviour considerably

- Relatively little manpower is required to implement the changes

- We (Cesnet) are ready to provide the outlined L&B extensions quickly (1-2 weeks)

- We (IT/CZ cluster) can demonstrate that the current problems are not intrinsic flaws of the WMS design