ProCKSI - Bug #20
Unsuccesful calculation of big data set

11/27/2007 09:58 AM - Paweł Widera

Status:	In Progress	**Start date:**	
Priority:	High	**Due date:**	
Assignee:	Anonymous	**% Done:**	0%
Category:	ProCKSI/server	**Estimated time:**	0.00 hour
Target version:	9.0		
Resolution:			

Description

Two thing has happened when I run the computations on a set of 269 chains:
- not all comparisons were made (31998 of 36315, see the vertical cut at the bottom of matrix): http://tinyurl.com/2gp5uk
- the archive file with all the results for Max-CMO and Dali-Lite is 0 bytes:
http://www.procksi.net/cgi-bin/archive.cgi?id_task=275&category=chains&source=ALL
http://www.procksi.net/cgi-bin/archive.cgi?id_task=276&category=chains&source=ALL

History

#1 - 05/22/2007 11:54 AM - Paweł Widera

When rerun on test server (request_id=9), archive file contains:

> Cannot exec tar: Argument list too long

Looks like different invocation of tar is needed (iterative addition of files and bzip and the end).

#2 - 07/18/2007 12:45 PM - Anonymous

(In r427) b - References #20

- Archives are generated on disk, not in memory (using 'tar')[[br]]
- Still, archive.cgi cannot handle larger datasets as the connection to the web browser gets lost

#3 - 07/18/2007 02:47 PM - Anonymous

- **Problems**[[br]]
  1. In the current approach, Archive::Tar::Wrapper COPIES all files to separate Temp Directory
  2. Generating the archive can take too long and the connection to the browser is lost.
- **Potential Solution for Prolem 1**[[br]]
  Add the Files sequentially directly to the archive without copying them using 'tar'. [[br]]
  But this would need multiple 'system' calls, each starting a new shell.
- **Potential Solution for Problem 2**[[br]]
  Divide functionality behind the "Download" button into two parts:
  1. *Generate Archive*: Generate an archive either as a background process or in the queue. Places archive into the task's home directory.
  2. *Download Archive*: Whenever the HTML page is refreshed, it's checked if an archive is available. If so, add this link to download the archive. Give the size of the archive.

#4 - 07/18/2007 02:49 PM - Anonymous

(In r428) e - References #20

Potential solution for problem 1 added

#5 - 07/18/2007 11:40 PM - Paweł Widera

It's good idea to prepare the archive asynchronously. It might be even more user friendly if some indication of progress would be presented to the user (like percentage of completeness) on every page refresh. You may consider using AJAX for that instead of meta-refresh.

#6 - 07/19/2007 09:46 AM - Anonymous

- AJAX technology is definatively a good candidate for many of the **frontend processes** that either have to handle loads of data or need to wait until one certain action is performed
- It might be even worth combining AJAX in the frontend with using a special queue with high priority for all **post-processing** tasks, e.g. preparing an archive for download

MAJOR PROBLEM:
Size of the results might be too big to download!

#7 - 11/27/2007 09:55 AM - Anonymous

- Status changed from New to In Progress