Mulgara - Bug #8

Temp directory management

07/11/2006 01:50 PM - brian -

Status:	New	**Start date:**	
Priority:	Normal	**Due date:**	
Assignee:	brian -	**% Done:**	0%
Category:	Mulgara	**Estimated time:**	0.00 hour
Target version:			
Resolution:			

Description

A research group has reported seeing Kowari fill up temp directories and fall over. This was on Solaris, but it might be a more general problem to solve. Nothing to reproduce it yet, but I just wanted to capture the experience to potentially investigate this issue moving forward.

The usage pattern was one big load and then mostly queries with the occasional insert.

History

#1 - 07/11/2006 02:39 PM - Paula Gearon

Queries create temporary files when constraint resolutions get too large to manage in memory. The result of a query may be small, but the results of individual constraints can be quite large, particularly when a lot of data has been loaded. So it may be these constraint resolution files at fault.

It would be worth testing if these files are being removed in a timely manner. I suggest adding an environment variable to override the high watermark for in-memory processing to a lower level, and then resolve several constraints which exceed this level.

The other thing to consider is how these files are being accessed. If they are managed through normal I/O calls, then the close() on the file should delete it (if if was opened as a "temporary" file). However, if they are being memory mapped, then we need to ensure that all references to the mapping are set to null. We can even use a System.gc() loop if we really need to make sure the file has gone (but that should be done as a last resort).

#2 - 07/11/2006 02:47 PM - brian -

It wouldn't have been the case on Solaris, but I know that there are problems with temp files not be deleted on Windows until the VM exits without a fair amount of nonsense. We should probably build some temp dir management tests into the suite (or extend any that are there) so we can easily catch these issues.

#3 - 09/06/2006 04:43 AM - Andrae Muys -

Some detailed bug reports would be useful, including the queries causing the trouble, and the count() attached to individual constraints.

There are some places where we might be performing distinct() or sort() more agressively then we strictly need to - these being the operations that generate temporary files. However most of these calls are no-ops as once an intermediate result is sorted, subsequent calls to sort/distinct involving that data tend to become no-ops.

To track this down, enable logging in [[HybridTuples]] and log both the file and it's provenance for each instance (watch for clones, naturally they share the same file).