

## Mulgara - Bug #8

### Temp directory management

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

<b>Status:</b>	New	<b>Start date:</b>	
<b>Priority:</b>	Normal	<b>Due date:</b>	
<b>Assignee:</b>	brian -	<b>% Done:</b>	0%
<b>Category:</b>	Mulgara	<b>Estimated time:</b>	0.00 hour
<b>Target version:</b>			
<b>Resolution:</b>			
<b>Description</b>			
<p>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.</p> <p>&lt;br/&gt;</p> <p>&lt;br/&gt;</p> <p>The usage pattern was one big load and then mostly queries with the occasional insert.</p>			

### 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.

<br/>

<br/>

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.

<br/>

<br/>

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 it 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 being deleted on Windows until the VM exits without a fair amount of nonsense. We should probably build some temp directory 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.

<br/>

<br/>

There are some places where we might be performing distinct() or sort() more aggressively than 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.

<br/>

<br/>

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