

Conversion Tools - Feature #2001

reporting v2.5

02/15/2013 07:25 PM - Greg Shah

Status:	Closed	Start date:	02/04/2013
Priority:	Normal	Due date:	
Assignee:	Greg Shah	% Done:	100%
Category:		Estimated time:	100.00 hours
Target version:	Conversion Support for Server Features	version:	
billable:	No		
vendor_id:	GCD		
Description			

History

#1 - 02/23/2013 12:20 PM - Greg Shah

- File *ges_upd20130215b.zip* added
- File *ges_upd20130215a.zip* added
- File *ges_upd20130215c.zip* added
- File *ges_upd20130206a.zip* added
- File *ges_upd20130207a.zip* added
- File *ges_upd20130208a.zip* added
- File *ges_upd20130212a.zip* added
- File *ges_upd20130213a.zip* added
- File *ges_upd20130214a.zip* added

This improvement to the reports is designed to massively improve performance. This was needed because on large projects (7+ million LOC), report runs can take 5 days or more. Eric profiled the reporting and the vast majority (90%?) of the time was spent in the reading of XML (loading from disk and parsing). This suggested the technical direction for this task.

The previous (v2.0) reporting approach had a PatternEngine run for each report. In each PatternEngine run, every AST in the project (or at least which are being included in reports) must be loaded/parsed and then it can be scanned for matches to the report condition. As you add reports, you linearly increase the number of times that the AST files have to be opened. On large projects this takes days with the current number of reports. And we plan to add reports. As a solution, I refactored all reports to process in a single PatternEngine run. This is the minimum XML processing time possible. To do this, I first had to move all the match processing and other report-level data into the database. Some of the match processing was already there, but I had to extend it. I also reworked all the report configuration into ReportDefinition objects that are registered with the worker, so the worker knows everything it is going on up front and can track some core information in this list. Everything else is in the DB now and the DB stores all reports in the same pass so it can get big. Of course, the DB had to be designed to differentiate a Statistic (now called MatchCategory) by report and so forth. The report generation had to be reworked to be fully driven off the ReportDefinition list and the database exclusively. All of the other data that was being stored on a per report basis has been eliminated or was moved into the ReportDefinition and out of the worker. Of course, the rule-sets (a new one called consolidated_reports.xml) and the ReportDriver had to have major changes to drive the process properly.

The result was a major improvement. It now takes 5-7 hours to run a project that otherwise would run in 13 hours. Please note that the OS reports that only 3 hrs 40 mins of that time was actually processing (CPU) time. So it is possible that the clock time of 5-7 hours could be even better.

I also really cleaned up the reports massively. Some of the reports had matching bugs and many had poorly defined categories (buckets). They are much improved now. Likewise, I added many reports where they were missing (based on a list of reports that I had to run ad-hoc to analyze the current project).

The final update (*ges_upd20130215c.zip*) passed regression testing (both conversion and runtime) and is checked in as bzip revision 10177.

If you are applying the zip file, the following files must be removed with this update (some of them were just moved in bzip but without bzip you must delete the files manually):

rules/reports/export_symbols.xml (deleted)
rules/reports/master_report* (deleted)
rules/reports/rpt_gen_file_targets.xml (deleted)
rules/reports/rpt_includes.xml (deleted)
rules/reports/rpt_lines_of_code.xml (deleted)

rules/reports/rpt_literals* (deleted)
rules/reports/rpt_template.xml (deleted)
src/com/goldencode/p2j/util/CopySpec.java (moved to pattern)
src/com/goldencode/p2j/util/Match.java (moved to pattern)
src/com/goldencode/p2j/util/Statistic (moved to pattern/MatchCategory)
src/com/goldencode/p2j/util/StatisticsHelper (moved to pattern)
src/com/goldencode/p2j/util/StatisticsWorker (moved to pattern/ReportWorker)
src/com/goldencode/p2j/util/SourceFileHelper.java (moved to pattern)

For your convenience, here is the command line:

rm rules/reports/export_symbols.xml rules/reports/master_report* rules/reports/rpt_gen_file_targets.xml rules/reports/rpt_includes.xml
rules/reports/rpt_lines_of_code.xml rules/reports/rpt_literals* rules/reports/rpt_template.xml src/com/goldencode/p2j/util/CopySpec.java
src/com/goldencode/p2j/util/Match.java src/com/goldencode/p2j/util/Stat* src/com/goldencode/p2j/util/SourceFileHelper.java

#2 - 03/29/2013 10:38 AM - Eric Faulhaber

- Status changed from WIP to Closed
- % Done changed from 0 to 100

#3 - 07/05/2013 05:04 PM - Greg Shah

- File ges_upd20130221c.zip added
- File ges_upd20130219a.zip added
- File ges_upd20130221a.zip added
- File ges_upd20130221b.zip added

The final update ges_upd20130221c.zip resolves issues related to unqualified constant names in TRPL rulesets. It is no longer possible to use an unqualified constant name. Constants must now ALWAYS be qualified with the worker's name (e.g. prog.kw_on instead of kw_on). This was checked into bzr as revision 10183.

The ges_upd20130221d.zip is a TIMCO Majic change that must be included with this update. See #2157.

These changes have passed regression testing (conversion only since the code changes don't affect runtime processing).

#4 - 11/16/2016 11:06 AM - Greg Shah

- Target version changed from Milestone 4 to Conversion Support for Server Features

Files			
ges_upd20130206a.zip	172 KB	02/23/2013	Greg Shah
ges_upd20130207a.zip	172 KB	02/23/2013	Greg Shah
ges_upd20130208a.zip	184 KB	02/23/2013	Greg Shah
ges_upd20130212a.zip	464 KB	02/23/2013	Greg Shah
ges_upd20130213a.zip	465 KB	02/23/2013	Greg Shah
ges_upd20130214a.zip	466 KB	02/23/2013	Greg Shah
ges_upd20130215a.zip	466 KB	02/23/2013	Greg Shah

ges_upd20130215b.zip	467 KB	02/23/2013	Greg Shah
ges_upd20130215c.zip	470 KB	02/23/2013	Greg Shah
ges_upd20130219a.zip	256 KB	07/05/2013	Greg Shah
ges_upd20130221a.zip	258 KB	07/05/2013	Greg Shah
ges_upd20130221b.zip	290 KB	07/05/2013	Greg Shah
ges_upd20130221c.zip	294 KB	07/05/2013	Greg Shah