

NOvA validation - Feature #14517

Reduce disk space usage?

11/15/2016 05:47 PM - Jeremy Wolcott

Status:	Closed	Start date:	11/15/2016
Priority:	Low	Due date:	
Assignee:		% Done:	0%
Category:		Estimated time:	0.00 hour
Target version:			
Description			
The output canvases written by MakeCanvases.py can occupy an awful lot of disk space (especially when there are multiple categories that all get multiplied together). Are there any ways of reducing it?			
One idea: storing gzipped versions of plots (will help for .json and .svg, at least). Not sure if this is possible, though -- would need to get Apache to serve these with a special directive to decompress them -- and even if that is doable, it might be more work (and Service Desk tickets) than it's worth.			
Probably need a better idea to move forward.			

Associated revisions

Revision 23056 - 11/18/2016 04:16 PM - Jeremy Wolcott

switch to using ROOT files instead of JSON for canvases. will (in some cases significantly) reduce disk space usage. (Redmine issue #14517)

History

#1 - 11/16/2016 03:25 PM - Christopher Backhouse

Are there json (or svg?) minifiers that help at all? Or are the files already pretty dense?

<https://feeding.cloud.geek.nz/posts/serving-pre-compressed-files-using/> for apache configuration?

#2 - 11/16/2016 03:33 PM - Jeremy Wolcott

Unfortunately you can't change the Apache headers sent with content without changing the directives in the master config file (not supported in .htaccess, I checked). Maybe it's worth fighting with SCD on that one if we can't find a better way, but I'm not keen on it.

If I had to guess, I'd say minifying the JSONs will maybe get us a factor of two, but not 10. Compression would be much better. I've discovered that the main culprit is actually SVGs that correspond to TH2s drawn with the "p" option, which I didn't realize was happening and have now disabled. I need to recheck the disk usage for realistic projects after that change.

#3 - 11/16/2016 03:34 PM - Jeremy Wolcott

Also, a minifier will of course slow down the processing even more. But maybe it's worth it, not sure.

#4 - 11/17/2016 03:21 PM - Jeremy Wolcott

Another, longer-term way to reduce the disk usage would be to offload all the logic for creating overlays and whatnot into the javascript. But that would basically be a rewrite of the underlying framework, and I'm sure I don't have the bandwidth for that right now. (Also I'm not sure how easy doing the statistical tests would be -- JSROOT doesn't implement anywhere near the whole ROOT histogram stack. At least, not yet.)

#5 - 11/17/2016 03:25 PM - Jeremy Wolcott

I also just discovered that JSROOT can read binary root files, which may be the easy-to-use compression scheme that we were looking for. I think that's probably the best path for me to look at, actually.

#6 - 11/18/2016 04:26 PM - Jeremy Wolcott

- Status changed from Feedback to Closed

Hopefully switching to ROOT file serialization ([novaart:r23056](https://redmine.cern.ch/issues/23056)) will be a good enough solution that I'm going to close this. (Certainly it will eliminate the >1 GB (!) 2D histogram problem that stemmed from having too many points and/or too many bins drawn in vector format.)