

GlideinWMS - Feature #24491

Multi Pronged Approach to Pilot Log Anonymization

05/28/2020 12:06 PM - Marco Mambelli

Status:	New	Start date:	05/28/2020
Priority:	Normal	Due date:	
Assignee:	Mirica Yancey	% Done:	0%
Category:	GlideinMonitor	Estimated time:	0.00 hour
Target version:	v3_7_1	Spent time:	0.00 hour
Stakeholders:			
Description			
GlideinMonitor is an application that allows viewing Glidein logs https://docs.google.com/document/d/1PqmQ_-JtqK472DT-Y0D08yU1m-IPeFvrp8n4x9f5QE/edit?usp=sharing https://github.com/glideinWMS/glideinmonitor			
A problem w/ the current monitor is that the logs may contain sensitive information Recognition of sensitive information and the subsequent anonymization should be added. This will use reversible and irreversible anonymization, done using different techniques like Named Entity Recognition, Regular Expression based pattern matching. GlideinMonitor's code includes an API for plug-ins to filter the logs that can be used or modified. The configuration of the anonymization software/plugin can happen via config files or a Web GUI.			
Steps will include:			
<ul style="list-style-type: none">• Research and evaluation of the state of the art for log anonymization• Code development (Python) and integration with the GlideinWMS and GlideinMonitor frameworks• Testing the solution in a distributed computing environment (Grids, Clouds, and Supercomputers) and presenting the results			
Some resources:			
<ul style="list-style-type: none">• https://aircloak.com/top-5-free-data-anonymization-tools/• https://airbus-cyber-security.com/reversible-log-anonymization-tool/• https://arxiv.org/abs/1706.04337• https://searchengineland.com/anonymizing-googles-server-log-data-hows-it-going-15036			