DAGMan Metrics Reporting

Purpose

Pegasus WMS is primarily a NSF funded project as part of the NSF SI2 track. The SI2 program focuses on robust, reliable, usable and sustainable software infrastructure that is critical to the CIF21 vision. As part of the requirements of being funded under this program, Pegasus WMS is required to gather usage statistics of Pegasus WMS and report it back to NSF in annual reports. The metrics will also enable us to improve our software as they will include errors encountered during the use of our software.

Associated Condor Ticket that covers the development of this feature in DAGMan

https://htcondor-wiki.cs.wisc.edu/index.cgi/tktview?tn=3532,4

DAGMan Metrics

All the metrics are sent in JSON format to a server at USC/ISI over HTTP.

http://metrics.pegasus.isi.edu/metrics

If condor team, wants a maintain a separate server or URL that is an option

Metrics can be turned off, on basis of an environment variable.

The proposal is to send metrics by Condor DAGMan whenever it exits

Proposed Metrics to be reported by DAGMan

Below are common metrics, that are shared with what pegasus-plan also reports currently

JSON KEY	DESCRIPTION
client	the name of the client (e.g "pegasus-plan")
version	the version of the client
type	type of data - "metrics"
start_time	start time of the client (in epoch seconds with millisecond precision)
end_time	end time of the client (in epoch seconds with millisecond precision)
duration	the duration of the condor_dagman
exitcode	the exitcode with which the dagman exits for a workflow
wf_uuid	the uuid of the executable workflow. It is generated by pegasus-plan at planning time. Can be null
root_wf_uuid	the uuid of the root workflow in case of hierarchal workflows. It is generated by pegasus-plan at planning time. Can be null

In addition, DAGMan we propose DAGMan send the following metrics

JSON KEY	DESCRIPTION
jobs	the number of vanilla jobs in the input DAG file
dag_jobs	the number of DAG jobs in the input DAG file i.e point to another DAG executed by another instance of DAGMan
total_jobs	the total number of jobs in the input DAG file
jobs_succeeded	the number of succeeded jobs/nodes in the workflow. don't count the DAG Nodes. (include retries?)
jobs_failed	the number of failed jobs/nodes in the workflow. don't count the DAG Nodes. (include retries?)
dag_jobs_succeeded	the number of DAG jobs that succeeded (include retries?)
dag_jobs_failed	the number of DAG jobs that failed. (include retries?)

total_jobs_run	the total number of jobs runs executed in a DAG. Should be equal to jobs_succeeded + jobs_failed + dag_jobs_succeeded + dag_jobs_failed