

JobSub - Feature #12598

Teach dagman to ignore exit statuses of jobs

05/11/2016 09:23 AM - Herbert Greenlee

Status:	Assigned	Start date:	05/11/2016
Priority:	Normal	Due date:	
Assignee:	Dennis Box	% Done:	0%
Category:		Estimated time:	0.00 hour
Target version:	v1.3.2	Spent time:	0.00 hour
Stakeholders:			

Description

Suppose I have a dag with a jobsub command that submits multiple parallel jobs, like this:

```
<parallel>
jobsub -N 100 ...
</parallel>
```

The way dagman works now, if any single job in the cluster returns with a nonzero exit status, the entire cluster will be killed by dagman. I would like this behavior to be changed so that each batch job is allowed to run to completion (i.e, teach dagman to ignore the exit statuses of jobs. Is there a way to do this already?).

I know that one can have a workaround by of reconfiguring the dag like this:

```
<parallel>
jobsub -N 1 ...
jobsub -N 1 ...
.
.
.
</parallel>
```

Besides being awkward, this workaround has the problem that such dags have a risk of timing out if the number of jobs is large.

History

#1 - 10/25/2016 04:03 PM - Dennis Box

- Status changed from New to Assigned
- Assignee set to Dennis Box
- Target version set to v1.2.4

#2 - 03/16/2017 10:00 AM - Dennis Box

Take a look at <beginjob> and <endjob> before closing this ticket. If they are easy to fix, do so, otherwise remove them from the documentation

#3 - 07/20/2017 09:24 PM - Dennis Box

- Target version changed from v1.2.4 to v1.2.5

Moving this to the next release.

Here is a discussion of the problem and progress towards a solution, from a status report dated 6/23/17:

Redmine issue [#12598](#), teach DAGs to continue after a node fails.

- Not as easy as I originally thought, for the following reasons:
- It is fairly easy to have an option that runs an 'exit 0' shell script after every DAG node, causing downstream nodes to run even if the job in the node fails.
- The --generate-email-summary option does this already.
 - requester Herb Greenlee wants a single node of his DAG to be multiple condor jobs all processing SAM data files.
 - The problem is that if any of the processes in the node fail, the whole node is marked as a failure and the DAG halts.

- If an 'exit 0' post script is appended to the node, the node is **not** marked as a failure and the downstream jobs run.
- This is still not what he wants: if any of the multiple jobs fail, condor aborts all the other jobs in the node immediately, and the data consumption of the entire set of processes halts.
- The solution is to run the multiple consuming processes inside its own DAG, with a cleanup stage as the final step of that DAG that runs no matter what the state of the individual processes.
 - This DAG needs to run inside other DAGs generated by the `jobsub_submit_dag` command.
 - I have this partially working
 - `jobsub_q` doesn't yet report its state correctly (NB since fixed)
 - some classad attributes are missing.

Other problems encountered, which make it prudent to move this feature to the next release:

- Generating multiple internal dags is not completely reliable, generation fails for some legal configurations of jobs
- I am worried about generating a circular graph of DAGs, I.E a cyclic graph or infinite loop.

#4 - 10/30/2017 11:43 AM - Dennis Box

- Target version changed from v1.2.5 to v1.3

#5 - 03/26/2019 11:14 AM - Dennis Box

- Target version changed from v1.3 to v1.3.2