

JobSub - Bug #6523

various problems with jobsub_submit.py

06/19/2014 10:03 AM - Dennis Box

Status:	Closed	Start date:	06/26/2014
Priority:	Normal	Due date:	07/30/2014
Assignee:	Dennis Box	% Done:	100%
Category:		Estimated time:	0.00 hour
Target version:	v0.4	Spent time:	16.00 hours
First Occurred:		Stakeholders:	
Occurs In:			

Description

On 6/19/14 9:45 AM, Joe Boyd wrote:

Hi Dennis,

Willis has the outputs of two jobs sitting in the directories noted below. He says the only difference between the two jobs is that one has parameters for his executable and the other doesn't (you can see that in the args line in the .cmd file). The problem is the out572 one fails because it doesn't find his script (you can see that in the .err file).

Look at the "Automatically generated by" lines in the _wrap file (including both below). Why does the one with parameters have the executable as "@/cdf/spool/willis/zcvms/runif.sh".

Does this make any sense to you or should I run some tests? I have in fact run many jobs that have parameters and it's fine. Is maybe the @ sign in Willis's parameters screwing up parsing or??????

One that doesn't work

1. Automatically generated by:
2. jobsub -l +JobsubJobId="\$(CLUSTER).\$(PROCESS)fifebatch2.fnal.gov" -l +Owner="willis" --resource-provides=usage_model=DEDICATED,OPPORTUNISTIC --OS=SL6 /cdf/spool/willis/zcvms/runif.sh 99 fngp5 willis@131.225.240.30 /cdf/spool/willis/zcvms *****

The one without params that does work

1. jobsub -l +JobsubJobId="\$(CLUSTER).\$(PROCESS)@fifebatch2.fnal.gov" -l +Owner="willis" --resource-provides=usage_model=DEDICATED,OPPORTUNISTIC --OS=SL6

/fife/local/scratch/uploads/cdf/willis/2014-06-17_124954.515138_3860/runif.sh

Thanks,

joe

On 06/19/2014 08:33 AM, Willis Sakumoto wrote:

Hi Joe:
The submit script and the outputs are on fcdflnx6,
/cdf/spool/willis/zcvms:

- a) outputs
ztemp/out564/ : no run params, it works
ztemp/out572/ : with run params, does not work
- b) submit script: ifsubmit.sh, the
' 99 \$RUNPAR \$USRHOST \$OUTBASE' is not there
for ztemp/out564/, and there for 'ztemp/out572/'

-- Willis

Subtasks:

Bug # 6561: regular expression in job.py still not right

Closed

Bug # 6563: dropbox storage area has lots of unnecessary duplicate files in oddly named...

Closed

History

#1 - 06/19/2014 02:46 PM - Dennis Box

Hi Willis,

What machine are you submitting from? I tried submitting from fcdflnx5 to make sure my workaround actually worked in the CDF environment, and it turns out jobsub_submit.py has authentication problems on that node see issue 6526

Thanks
Dennis

#2 - 06/26/2014 04:07 PM - Dennis Box

Willis' problem as described stems from one of the input parameters to runif.sh :

```
jobsub --resource-provides=usage_model=DEDICATED,OPPORTUNISTIC --OS=SL6  
file:///runif.sh 99 fngp5 willis@131.225.240.30 /cdf/spool/willis/zcvdfs
```

The email address willis@131.225.240.30 has an " in it. job.py has a regular expression that cues on " to change the path of runif.sh from the clients location to the servers. It breaks if there is an email address as an argument after runif.sh

As a workaround I advised willis to omit the '@' in his input parameters and co-join 'willis' and 131.225.240.30 programmatically in his script runif.sh.

Looking deeper this regex is a consequence of <https://cdcv.s.fnal.gov/redmine/issues/5813>. job.py originally had a regex that broke when someone submitted a job with an email address **before** the executable. We (and by we I mean me) stupidly changed the regular expression to cue on the last '@' in the arguments which broke when Willis put his email address as an argument after the executable.

I found other problems looking at the interplay of --tarfile, file://some_file, dropbox://some_file, I am making subtask tickets:

subtask 1): fix regular expression so its happy with email addresses before or after executable. Fixing regex so it never breaks when users discover a new way to put '@' in an argument may be harder, I'm not as imaginative as some users. I have tested one that seems pretty bullet proof but one never knows .

While looking for a workaround for Willis I realized these are needed:

subtask 2): dropbox:///some_file needs to put servers /path/to/some_file in the transfer_input_files list.

subtask 3): the dropbox storage area has lots of unnecessary duplicate files in oddly named directories. Potential for lots of unnecessary uploading to the server, and a operational headache when the dropbox area fills up with duplicates. The dropbox area should have subdirectories named after a hex digest of the uploaded file, if file already lives on dropbox server don't upload it again. This has the added benefit of ensuring that the file was not corrupted when uploaded by checking clients digest against servers digest.

#3 - 06/26/2014 05:47 PM - Dennis Box

- Subject changed from *--tarfile problems with jobsub_submit.py* to *various problems with jobsub_submit.py*

- Status changed from *New* to *Resolved*

I changed the title of this issue as I confused this ticket with other questions Willis had about the `--tarfile` option. The `--tarfile` option never worked very well and is not needed given that the `-f` option with `ifdh` and the `dropbox://` option do the same thing only better. Willis cannot use `ifdh` and the `-f` and `-d` options as these require `cdf` lockfiles which do not exist and may never. He will be able to use `dropbox://` with the release of 0.4

#4 - 07/28/2014 10:45 AM - Parag Mhashilkar

- Status changed from *Resolved* to *Closed*