

## GlideinWMS - Feature #3150

### Speeding up the FE matchmaking - pre-clustering

11/29/2012 02:50 PM - Igor Sfiligoi

<b>Status:</b>	Closed	<b>Start date:</b>	12/04/2012
<b>Priority:</b>	Normal	<b>Due date:</b>	
<b>Assignee:</b>	Igor Sfiligoi	<b>% Done:</b>	50%
<b>Category:</b>	Frontend	<b>Estimated time:</b>	0.00 hour
<b>Target version:</b>	v2_7	<b>Spent time:</b>	1.00 hour
<b>Stakeholders:</b>			
<b>Description</b>			
This is a proposal for speeding up the matchmaking in the FE.  Currently we are matching every job independently. Instead, the proposal is to pre-cluster them, get the count, and just match the clusters. (where cluster == jobs with the same values for the attributes used in matchmaking)			
<b>Subtasks:</b>			
Idea # 3160: Reducing the memory footprint of the FE with clustering			<b>New</b>
Feature # 3196: Master: Speeding up the FE matchmaking - pre-clustering			<b>Closed</b>

#### History

##### #1 - 11/29/2012 02:51 PM - Igor Sfiligoi

Speeding up the FE matchmaking is a high priority task.

CMS AnaOps FE is taking 40 mins to match the jobs as I write this!

##### #2 - 12/04/2012 10:49 AM - Burt Holzman

I looked at this -- I think we actually did implement condor-style autoclustering in 971470670. We should consider clustering the glidename too, but how much time is saved will depend on the match expression. Igor's going to double-check me.

##### #3 - 12/04/2012 01:12 PM - Douglas Strain

- Assignee changed from Burt Holzman to Douglas Strain

##### #4 - 12/04/2012 02:36 PM - Igor Sfiligoi

- File `glideinFrontendLib.py.patch` added

I think I have found and fixed the problem.

My last patch was partial; only half of the function was clusterized :(

I have now finished the work, and clusterized what is missing.  
Will commit to git tomorrow (strict firewall right now), but attaching the patch.

PS: I have applied the patch to the CMS AnaOps FE, and the matching time went from almost an hour to about 2mins ;)

##### #5 - 12/04/2012 03:06 PM - Parag Mhashilkar

- Target version changed from v2\_7\_x to 293

##### #6 - 12/10/2012 01:39 PM - Douglas Strain

- Status changed from New to Feedback

- Assignee changed from Douglas Strain to Igor Sfiligoi

If I understand this correctly, the code looks correct. The only functional comment is why multiple by `nr_schedds` on lines 275,280? I'm not sure I understand what's going on there.

However, this code badly needs some code comments, since it is extremely confusing and difficult to understand.

I think there should be some more comments, such as what data types each of these variables with examples. I think once that is done, it is okay to

merge.

**#7 - 12/17/2012 11:04 AM - Burt Holzman**

Igor,

This request from Doug seems reasonable. Could you please comment the code so we can merge this?

- B

**#8 - 12/17/2012 02:37 PM - Igor Sfiligoi**

I sympathize with the request for more comments, but don't have time right now to re-comment a large portion of the code. If you have a request for a specific section of the patch that needs to be commented, i can attempt that... but not much more in the short term, sorry.

**#9 - 12/17/2012 03:16 PM - Parag Mhashilkar**

- *Tracker changed from Idea to Feature*
- *Target version changed from 293 to v2\_7\_x*

In the meeting we decided to push it off to future release until code was documented.

**#10 - 12/17/2012 05:37 PM - Igor Sfiligoi**

I have added a comment that describes the indexing algorithm that confused Doug above.

**#11 - 12/18/2012 11:49 AM - Douglas Strain**

I must be in the Christmas spirit. I have gone in and commented most of this code.

Igor, can you do the following:

- 1) Quickly review my comments for correctness
- 2) Anywhere I have a "???", add in the relevant details.
  - 2a) Fill in "???" for the undocumented parameters/types in the epydoc docstring below the function def
  - 2b) Towards the middle of the function, I sort of lost track of the logic. All the manipulation of data structures lost me. Why do we need `list_of_all_jobs` which turns into `(outvals_cl, jrange_cl)`. Why is it needed in that format? Can you put in a comment or two elaborating what's going on here? Thanks.

Hopefully, this should be less time-consuming than commenting all the code. Just fill in the blanks. Once this is done, I think we can merge it.

Doug Strain

**#12 - 12/18/2012 12:33 PM - Igor Sfiligoi**

Updated the comments where needed. The rest looks good.

**#13 - 12/18/2012 01:17 PM - Douglas Strain**

It looks good to me. I think we are ready to merge.

**#14 - 12/19/2012 02:04 PM - Parag Mhashilkar**

- *Status changed from Feedback to Resolved*
- *Target version changed from v2\_7\_x to 293*

Closing this one. Opened a new ticket for merging the changes to master

**#15 - 03/18/2013 01:20 PM - Parag Mhashilkar**

- *Target version changed from 293 to v2\_7*

**#16 - 03/18/2013 01:56 PM - Parag Mhashilkar**

- *Status changed from Resolved to Closed*

**Files**

---

glideinFrontendLib.py.patch	1.93 KB	12/04/2012	Igor Sfiligoi
-----------------------------	---------	------------	---------------