Subject: Fwd: Auto-tuning user jobs
From: Stefano Belforte <stefano.belforte@cern.ch>
Date: 20/04/17 01:24
CC: <stefano.belforte@cern.ch>

— Auto-tuning user jobs.eml

Subject: Auto-tuning user jobs
From: Brian Bockelman <bbockelm@cse.unl.edu>
Date: 09/04/17 02:57
To: Jean-Roch Vlimant <jean-roch.vlimant@cern.ch>, Stefano Belforte <stefano.belforte@cern.ch>

Hi Jean-Roch, Stefano,

Once there's a certain level of available statistics, we automatically tune production jobs' requested wall time to be equal to the 95th percentile of the runtimes for that workflow.

Can we do the same for analysis jobs?

I.e., have a formula along the lines of:

MaxWallTimeMins = min(original request, max(all finished jobs) + 2 hours)

Thoughts? Would really improve the utilization of the pool.

Brian

— Re: Auto-tuning user jobs.eml

Subject: Re: Auto-tuning user jobs
From: Stefano Belforte <stefano.belforte@cern.ch>
Date: 09/04/17 14:55
To: Brian Bockelman <bbockelm@cse.unl.edu>, Jean-Roch Vlimant <jean-roch.vlimant@cern.ch>
CC: <stefano.belforte@cern.ch>

interesting idea. How do we actully do it ?
jobrouter or massinve condor_qedit ?

I suspect this came up years ago talking with Igor,
but largish scale condor_qedit was known to kill the pool,
at that time.

Stefano

On 09/04/17 02:57, Brian Bockelman wrote:

> Hi Jean-Roch, Stefano,
>
> Once there's a certain level of available statistics, we automatically tune production jobs' requested wall time to be equal to the 95th percentile of the runtimes for that workflow.

> Can we do the same for analysis jobs?
>
> I.e., have a formula along the lines of:
>
> MaxWallTimeMins = min(original request, max(all finished jobs) + 2 hours)
>
> Thoughts? Would really improve the utilization of the pool.
>
> Brian

— Re: Auto-tuning user jobs.eml

Subject: Re: Auto-tuning user jobs
From: Brian Bockelman <bbockelm@cse.unl.edu>
Date: 09/04/17 16:56
To: Stefano Belforte <stefano.belforte@cern.ch>
CC: Jean-Roch Vlimant <jean-roch.vlimant@cern.ch>

This would be done with the jobrouter - we do the same for the production systems. Jobrouter is known to scale well for this.

Put in a small PR on CRABServer to differentiate between wall time for matchmaking and wall time for killing jobs. That way, we can be more accurate without potentially affecting users.

Brian

Sent from my iPhone

> On Apr 9, 2017, at 7:55 AM, Stefano Belforte <stefano.belforte@cern.ch> wrote:
>
> interesting idea. How do we actully do it ?
> jobrouter or massinve condor_qedit ?
>
> I suspect this came up years ago talking with Igor,
> but largish scale condor_qedit was known to kill the pool,
> at that time.
>
> Stefano
>
>> On 09/04/17 02:57, Brian Bockelman wrote:
>> Hi Jean-Roch, Stefano,
>>
>> Once there's a certain level of available statistics, we automatically tune production jobs' requested wall time to be equal to the 95th percentile of the runtimes for that workflow.
>>
>> Can we do the same for analysis jobs?
>>
>> I.e., have a formula along the lines of:
>>
>> MaxWallTimeMins = min(original request, max(all finished jobs) + 2 hours)
>>
>> Thoughts? Would really improve the utilization of the pool.
>>
>> Brian

— Re: Auto-tuning user jobs.eml —

Subject: Re: Auto-tuning user jobs
From: Stefano Belforte <stefano.belforte@cern.ch>
Date: 10/04/17 09:29
To: Brian Bockelman <bbockelm@cse.unl.edu>
CC: Jean-Roch Vlimant <jean-roch.vlimant@cern.ch>

On 04/09/2017 04:56 PM, Brian Bockelman wrote:

> This would be done with the jobrouter - we do the same for the production systems. Jobrouter is known to scale well for this.

we have a JobRrouter on each CRAB schedd's to allow submission to LPC.
https://gitlab.cern.ch/ai/it-puppet-hostgroup-vocmsglidein/blob/master/code/templates/modules/condor/90_cmslpc_jobrouter.config
I know nothing about its configuration, usage etc. Should we get SI to deal
with this ? Crab Cat-A operator, Todor, has to be on top eventually,
but needs traning. Who is at CERN who can guide him on this ?

Is JeanRoch in the thread so he can provide the code used in production
as a start ?

> Put in a small PR on CRABServer to differentiate between wall time for matchmaking and wall time for killing jobs. That way, we can be more accurate without potentially affecting users.

I saw it, thanks. That's really helpful.
Stefano

> Brian
>
> Sent from my iPhone
>
>> On Apr 9, 2017, at 7:55 AM, Stefano Belforte <stefano.belforte@cern.ch> wrote:
>>
>> interesting idea. How do we actully do it ?
>> jobrouter or massinve condor_qedit ?
>>
>> I suspect this came up years ago talking with Igor,
>> but largish scale condor_qedit was known to kill the pool,
>> at that time.
>>
>> Stefano
>>
>>> On 09/04/17 02:57, Brian Bockelman wrote:
>>> Hi Jean-Roch, Stefano,
>>>
>>> Once there's a certain level of available statistics, we automatically tune production jobs' requested wall time to be equal to the 95th percentile of the runtimes for that workflow.
>>>
>>> Can we do the same for analysis jobs?
>>>
>>> I.e., have a formula along the lines of:
>>>
>>> MaxWallTimeMins = min(original request, max(all finished jobs) + 2 hours)
>>>
>>> Thoughts? Would really improve the utilization of the pool.

>>> Brian

— Re: Auto-tuning user jobs.eml —

Subject: Re: Auto-tuning user jobs
From: Jean-Roch Vlimant <jean-roch.vlimant@cern.ch>
Date: 10/04/17 10:59
To: Stefano Belforte <stefano.belforte@cern.ch>, Brian Bockelman <bbockelm@cse.unl.edu>

Hi,

the code that is ran in JobRouter to create the production rules is

https://github.com/CMSCompOps/WmAgentScripts/blob/master/Unified/go_condor.py

which reads from

https://cmst2.web.cern.ch/cmst2/unified/equalizor.json

I am available to discuss details if needed.
The above does not include code to read monitor data and make the 95% percentile though.

Cheers
Jean-Roch

On 4/10/17 9:29 AM, Stefano Belforte wrote:

> On 04/09/2017 04:56 PM, Brian Bockelman wrote:
>
>> This would be done with the jobrouter - we do the same for the production systems. Jobrouter is known to scale well for this.
>
> we have a JobRrouter on each CRAB schedd's to allow submission to LPC.
> https://gitlab.cern.ch/ai/it-puppet-hostgroup-vocmsglidein/blob/master/code/templates/modules/condor/90_cmslpc_jobrouter.config
>
> I know nothing about its configuration, usage etc. Should we get SI to deal with this ? Crab Cat-A operator, Todor, has to be on top eventually, but needs traning. Who is at CERN who can guide him on this ?
>
> Is JeanRoch in the thread so he can provide the code used in production as a start ?
>
>> Put in a small PR on CRABServer to differentiate between wall time for matchmaking and wall time for killing jobs. That way, we can be more accurate without potentially affecting users.
>
> I saw it, thanks. That's really helpful.
> Stefano
>
>> Brian
>>
>> Sent from my iPhone
>>
>>> On Apr 9, 2017, at 7:55 AM, Stefano Belforte

>>> <stefano.belforte@cern.ch> wrote:
>>>
>>> interesting idea. How do we actully do it ?
>>> jobrouter or massinve condor_qedit ?
>>>
>>> I suspect this came up years ago talking with Igor,
>>> but largish scale condor_qedit was known to kill the pool,
>>> at that time.
>>>
>>> Stefano
>>>
>>>> On 09/04/17 02:57, Brian Bockelman wrote:
>>>> Hi Jean-Roch, Stefano,
>>>>
>>>> Once there's a certain level of available statistics, we
>>>> automatically tune production jobs' requested wall time to be equal
>>>> to the 95th percentile of the runtimes for that workflow.
>>>>
>>>> Can we do the same for analysis jobs?
>>>>
>>>> I.e., have a formula along the lines of:
>>>>
>>>> MaxWallTimeMins = min(original request, max(all finished jobs) + 2
>>>> hours)
>>>>
>>>> Thoughts? Would really improve the utilization of the pool.
>>>>
>>>> Brian

```
--
-----------------------------------------------------------
 Dr. Jean-Roch Vlimant. California Institute of Technology
     CERN : 32-3A-015  +412276 70106  +417541 15353
-----------------------------------------------------------
```

Re: Auto-tuning user jobs.eml

Subject: Re: Auto-tuning user jobs
From: Brian Bockelman <bbockelm@cse.unl.edu>
Date: 10/04/17 17:00
To: Stefano Belforte <stefano.belforte@cern.ch>
CC: Jean-Roch Vlimant <jean-roch.vlimant@cern.ch>

> On Apr 10, 2017, at 2:29 AM, Stefano Belforte <stefano.belforte@cern.ch> wrote:
>
>
> On 04/09/2017 04:56 PM, Brian Bockelman wrote:
>> This would be done with the jobrouter - we do the same for the production systems.
>> Jobrouter is known to scale well for this.
>
> we have a JobRrouter on each CRAB schedd's to allow submission to LPC.
> https://gitlab.cern.ch/ai/it-puppet-hostgroup-vocmsglidein/blob/master/code/templates/modules/condor/90_cmslpc_jobrouter.config
> I know nothing about its configuration, usage etc. Should we get SI to deal
> with this ? Crab Cat-A operator, Todor, has to be on top eventually,
> but needs traning. Who is at CERN who can guide him on this ?

Likely some combination of Jean-Roch and Justas.

I can help if someone could give my CERN account (bbockelm) permission to view this repository!

> Is JeanRoch in the thread so he can provide the code used in production as a start ?

Yup! Basically, the information flow is:

gwmsmon (Justas; provides monitoring info) -> Unified (Jean-Roch; digests monitoring info into a per-task policy) -> JobRouter (Brian; converts per-task policy into set of rules for altering jobs)

Given there is no equivalent of Unified here, I would suggest:

gwmsmon (Justas; provides monitoring info per task) -> JobRouter (Todor? converts monitoring to JobRouter rules)

> > Put in a small PR on CRABServer to differentiate between wall time for matchmaking and wall time for killing jobs. That way, we can be more accurate without potentially affecting users.
>
> I saw it, thanks. That's really helpful.

Very small first step!

(Although, admittedly not a particularly long road!)

Brian

— Re: Auto-tuning user jobs.eml

Subject: Re: Auto-tuning user jobs
From: Stefano Belforte <stefano.belforte@cern.ch>
Date: 10/04/17 17:20
To: Todor Trendafilov Ivanov <todor.trendafilov.ivanov@cern.ch>
CC: Brian Bockelman <bbockelm@cse.unl.edu>

Todor, can you please make sure that Brian can read our pupper repository ?
I am surprised that it is not configured with read access to everybody in CMS, actually .

Also.. please note that we are discussing here how to give you yet another things to do. I know that new things pile up faster then you get old ones out of the way, it is the same for everybody I fear.

So let me know what you think about getting involved with JobRouter configuration in CRAB schedd's. I am afraid that it fits very well your job description 🙂 And since we run it on our schedd's anyhow, you have to learn how it works in any case !

Stefano

On 04/10/2017 05:00 PM, Brian Bockelman wrote:

> > On Apr 10, 2017, at 2:29 AM, Stefano Belforte <stefano.belforte@cern.ch> wrote:

> On 04/09/2017 04:56 PM, Brian Bockelman wrote:
>> This would be done with the jobrouter - we do the same for the production systems. Jobrouter is known to scale well for this.
>
> we have a JobRrouter on each CRAB schedd's to allow submission to LPC.
> https://gitlab.cern.ch/ai/it-puppet-hostgroup-vocmsglidein/blob/master/code/templates/modules/condor/90_cmslpc_jobrouter.config
> I know nothing about its configuration, usage etc. Should we get SI to deal
> with this ? Crab Cat-A operator, Todor, has to be on top eventually,
> but needs traning. Who is at CERN who can guide him on this ?

Likely some combination of Jean-Roch and Justas.

I can help if someone could give my CERN account (bbockelm) permission to view this repository!

> Is JeanRoch in the thread so he can provide the code used in production
> as a start ?

Yup! Basically, the information flow is:

gwmsmon (Justas; provides monitoring info) -> Unified (Jean-Roch; digests monitoring info into a per-task policy) -> JobRouter (Brian; converts per-task policy into set of rules for altering jobs)

Given there is no equivalent of Unified here, I would suggest:

gwmsmon (Justas; provides monitoring info per task) -> JobRouter (Todor? converts monitoring to JobRouter rules)

>> Put in a small PR on CRABServer to differentiate between wall time for matchmaking and wall time for killing jobs. That way, we can be more accurate without potentially affecting users.
>
> I saw it, thanks. That's really helpful.

Very small first step!

(Although, admittedly not a particularly long road!)

Brian

— Re: Auto-tuning user jobs.eml —

Subject: RE: Auto-tuning user jobs
From: Todor Trendafilov Ivanov <todor.trendafilov.ivanov@cern.ch>
Date: 10/04/17 17:33
To: Stefano Belforte <Stefano.Belforte@cern.ch>
CC: "bbockelm@cse.unl.edu" <bbockelm@cse.unl.edu>

For sure I will get the assignment 🙂 even it was not on my job description.
This puppet repository that we have must be available for Brian. If it is not we will
have to check with Daniel tomorrow. For Diego it is available for sure and I remember
that we had some discussion about this lpc enabling on our site few days ago. It was

related to the way it should be enabled for the testing machine that I am using because I had to debug a little bit the reason why that think was not working for the vocms0115 and found that we need a firewall opened from FNAL side but I just disabled the functionality for the moment in that machine because my work was not related to that directly. So that's all I know for this mechanism for the moment. I definitely should get more deeply in touch with that one as it is installed on our machines.

Regards,
Todor

From: Stefano Belforte
Sent: 10 April 2017 17:20
To: Todor Trendafilov Ivanov
Cc: bbockelm@cse.unl.edu
Subject: Re: Auto-tuning user jobs

Todor, can you please make sure that Brian can read our pupper repository ?
I am surprised that it is not configured with read access to everybody in CMS, actually.

Also.. please note that we are discussing here how to give you yet another things to do. I know that new things pile up faster then you get old ones out of the way, it is the same for everybody I fear.

So let me know what you think about getting involved with JobRouter configuration in CRAB schedd's. I am afraid that it fits very well your job description 🙂 And since we run it on our schedd's anyhow, you have to learn how it works in any case !

Stefano

On 04/10/2017 05:00 PM, Brian Bockelman wrote:

> > On Apr 10, 2017, at 2:29 AM, Stefano Belforte <stefano.belforte@cern.ch> wrote:
> >
> > On 04/09/2017 04:56 PM, Brian Bockelman wrote:
> >
> > > This would be done with the jobrouter - we do the same for the production systems. Jobrouter is known to scale well for this.
> >
> > we have a JobRouter on each CRAB schedd's to allow submission to LPC.
> > https://gitlab.cern.ch/ai/it-puppet-hostgroup-vocmsglidein/blob/master/code/templates/modules/condor/90_cmslpc_jobrouter.config
> > I know nothing about its configuration, usage etc. Should we get SI to deal with this ? Crab Cat-A operator, Todor, has to be on top eventually, but needs traning. Who is at CERN who can guide him on this ?
>
> Likely some combination of Jean-Roch and Justas.
>
> I can help if someone could give my CERN account (bbockelm) permission to view this repository!
>
> > Is JeanRoch in the thread so he can provide the code used in production as a start ?
>
> Yup! Basically, the information flow is:
>
> gwmsmon (Justas; provides monitoring info) -> Unified (Jean-Roch; digests monitoring info into a per-task policy) -> JobRouter (Brian; converts per-task policy into set of rules for altering jobs)

> Given there is no equivalent of Unified here, I would suggest:
>
> gwmsmon (Justas; provides monitoring info per task) -> JobRouter (Todor? converts monitoring to JobRouter rules)
>
> > > Put in a small PR on CRABServer to differentiate between wall time for matchmaking and wall time for killing jobs. That way, we can be more accurate without potentially affecting users.
> >
> > I saw it, thanks. That's really helpful.
>
> Very small first step!
>
> (Although, admittedly not a particularly long road!)
>
> Brian

— Re: Auto-tuning user jobs.eml

Subject: RE: Auto-tuning user jobs
From: Todor Trendafilov Ivanov <todor.trendafilov.ivanov@cern.ch>
Date: 10/04/17 20:49
To: Stefano Belforte <Stefano.Belforte@cern.ch>
CC: "bbockelm@cse.unl.edu" <bbockelm@cse.unl.edu>, Daniel Valbuena Sosa <daniel.valbuena.sosa@cern.ch>

Hi Brian,

I am not 100% sure but I think that in order to have access to the it-puppet-hostgroup-vocms* repositories in gitlab you should add yourself to the following e-groups:
ai-admins
cloud-infrastructure-users

or at least Daniel can add you or ask for your access there (that's why I am adding him as a CC in that list).
Daniel please correct me if I am wrong. I do not remember if I was able to join these e-groups by myself or I had to ask for access.

Regards,
Todor

— Re: Auto-tuning user jobs.eml

Subject: Re: Auto-tuning user jobs
From: Daniel Valbuena Sosa <daniel.valbuena.sosa@cern.ch>
Date: 11/04/17 09:50
To: Todor Trendafilov Ivanov <todor.trendafilov.ivanov@cern.ch>
CC: Stefano Belforte <Stefano.Belforte@cern.ch>, "bbockelm@cse.unl.edu" <bbockelm@cse.unl.edu>

Dear All,

I added Brian in https://gitlab.cern.ch/ai/it-puppet-hostgroup-vocms project.
Please let me know if you need additional permissions.

Best,

Daniel.

On Mon, Apr 10, 2017 at 8:49 PM, Todor Trendafilov Ivanov
<todor.trendafilov.ivanov@cern.ch> wrote:

> Hi Brian,
>
> I am not 100% sure but I think that in order to have access to the
> it-puppet-hostgroup-vocms* repositories in gitlab you should add yourself to
> the following e-groups:
> ai-admins
> cloud-infrastructure-users
>
> or at least Daniel can add you or ask for your access there (that's why I am
> adding him as a CC in that list).
> Daniel please correct me if I am wrong. I do not remember if I was able to
> join these e-groups by myself or I had to ask for access.
>
> Regards,
> Todor

— Re: Auto-tuning user jobs.eml

Subject: Re: Auto-tuning user jobs
From: Stefano Belforte <stefano.belforte@cern.ch>
Date: 11/04/17 10:01
To: Daniel Valbuena Sosa <daniel.valbuena.sosa@cern.ch>, "Todor
Trendafilov Ivanov" <todor.trendafilov.ivanov@cern.ch>
CC: "bbockelm@cse.unl.edu" <bbockelm@cse.unl.edu>

sorry, the mail I meant to sent yesterday was still an unsent draft.
I had manged to add Brian with role "reporter"
to AIGROUP-cms-service-glideinwms-gitlab
I hope I did not do anything wrong, it is anyhow odd that
there are apparently 445 Members with access to
it-puppet-hostgroup-vocmsglidein, many as developer

That group is where I was lead from
https://gitlab.cern.ch/ai/it-puppet-hostgroup-vocmsglidein/edit
It seems visibility of the ai / it-puppet-hostgroup-vocmsglidein
is set to PRIVATE, which IMHO is too restrictive.

I have no idea of the relation between
https://gitlab.cern.ch/ai/it-puppet-hostgroup-vocmsglidein

and
https://gitlab.cern.ch/ai/it-puppet-hostgroup-vocms

I am really confused, I must confess !

Stefano

On 04/11/2017 09:50 AM, Daniel Valbuena Sosa wrote:

> Dear All,
>
> I added Brian in https://gitlab.cern.ch/ai/it-puppet-hostgroup-vocms project.
> Please let me know if you need additional permissions.
>
> Best,
>
> Daniel.
>
> On Mon, Apr 10, 2017 at 8:49 PM, Todor Trendafilov Ivanov <todor.trendafilov.ivanov@cern.ch
> <mailto:todor.trendafilov.ivanov@cern.ch>> wrote:
>
> > Hi Brian,
> >
> > I am not 100% sure but I think that in order to have access to the it-puppet-hostgroup-vocms*
> > repositories in gitlab you should add yourself to the following e-groups:
> > ai-admins
> > cloud-infrastructure-users
> >
> > or at least Daniel can add you or ask for your access there (that's why I am adding him as a CC
> > in that list).
> > Daniel please correct me if I am wrong. I do not remember if I was able to join these e-groups
> > by myself or I had to ask for access.
> >
> > Regards,
> > Todor

— Re: Auto-tuning user jobs.eml —

Subject: Re: Auto-tuning user jobs
From: Brian Bockelman <bbockelm@cse.unl.edu>
Date: 19/04/17 05:46
To: Stefano Belforte <stefano.belforte@cern.ch>
CC: Jean-Roch Vlimant <jean-roch.vlimant@cern.ch>, Todor Trendafilov Ivanov <todor.trendafilov.ivanov@cern.ch>

Hi,

I'd like to proceed along this route. How can we best move things forward?

Thanks,

Brian

Sent from my iPhone

> On Apr 10, 2017, at 10:00 AM, Brian Bockelman <bbockelm@cse.unl.edu> wrote:
>
> > On Apr 10, 2017, at 2:29 AM, Stefano Belforte <stefano.belforte@cern.ch> wrote:
> >
> > > On 04/09/2017 04:56 PM, Brian Bockelman wrote:
> > > This would be done with the jobrouter - we do the same for the production systems.
> > > Jobrouter is known to scale well for this.
> >
> > we have a JobRouter on each CRAB schedd's to allow submission to LPC.
> > https://gitlab.cern.ch/ai/it-puppet-hostgroup-vocmsglidein/blob/master/code/templates/modules/condor/90_cmslpc_jobrouter.config
> > I know nothing about its configuration, usage etc. Should we get SI to deal
> > with this ? Crab Cat-A operator, Todor, has to be on top eventually,
> > but needs traning. Who is at CERN who can guide him on this ?
>
> Likely some combination of Jean-Roch and Justas.
>
> I can help if someone could give my CERN account (bbockelm) permission to view this repository!
>
> > Is JeanRoch in the thread so he can provide the code used in production
> > as a start ?
>
> Yup! Basically, the information flow is:
>
> gwmsmon (Justas; provides monitoring info) -> Unified (Jean-Roch; digests monitoring info into a per-task policy) -> JobRouter (Brian; converts per-task policy into set of rules for altering jobs)
>
> Given there is no equivalent of Unified here, I would suggest:
>
> gwmsmon (Justas; provides monitoring info per task) -> JobRouter (Todor? converts monitoring to JobRouter rules)
>
> > > Put in a small PR on CRABServer to differentiate between wall time for matchmaking
> > > and wall time for killing jobs. That way, we can be more accurate without
> > > potentially affecting users.
> >
> > I saw it, thanks. That's really helpful.
>
> Very small first step!
>
> (Although, admittedly not a particularly long road!)
>
> Brian

— Re: Auto-tuning user jobs.eml

Subject: Re: Auto-tuning user jobs
From: Stefano Belforte <stefano.belforte@cern.ch>
Date: 19/04/17 07:35
To: Brian Bockelman <bbockelm@cse.unl.edu>

CC: <stefano.belforte@cern.ch>, Jean-Roch Vlimant <jean-roch.vlimant@cern.ch>, Todor Trendafilov Ivanov <todor.trendafilov.ivanov@cern.ch>

Todor should be able to start working on this starting this week or next one at latest. Hopefully he and Diego Davila can start by understanding how current JobRouter-for-LPC works. I am a bit lost at where we are on the gwmsmon part, although it came to my mined that we can ask Emilis to provide the up-to-date timing info as a file on the schedd spool director via the TaskProcess. That TP npw runs every 5 min and parses condor log for the task to update on jobs status. That would make the whole thing local to the schedd, while the gwmsmon approach has more generality.

Stefano

On 19/04/17 05:46, Brian Bockelman wrote:

> Hi,
>
> I'd like to proceed along this route. How can we best move things forward?
>
> Thanks,
>
> Brian
>
> Sent from my iPhone
>
>> On Apr 10, 2017, at 10:00 AM, Brian Bockelman <bbockelm@cse.unl.edu> wrote:
>>
>>> On Apr 10, 2017, at 2:29 AM, Stefano Belforte <stefano.belforte@cern.ch> wrote:
>>>
>>>> On 04/09/2017 04:56 PM, Brian Bockelman wrote:
>>>> This would be done with the jobrouter - we do the same for the production systems. Jobrouter is known to scale well for this.
>>>
>>> we have a JobRrouter on each CRAB schedd's to allow submission to LPC.
>>> https://gitlab.cern.ch/ai/it-puppet-hostgroup-vocmsglidein/blob/master/code/templates/modules/condor/90_cmslpc_jobrouter.config
>>> I know nothing about its configuration, usage etc. Should we get SI to deal with this ? Crab Cat-A operator, Todor, has to be on top eventually, but needs traning. Who is at CERN who can guide him on this ?
>>
>> Likely some combination of Jean-Roch and Justas.
>>
>> I can help if someone could give my CERN account (bbockelm) permission to view this repository!
>>
>>> Is JeanRoch in the thread so he can provide the code used in production as a start ?
>>
>> Yup! Basically, the information flow is:
>>
>> gwmsmon (Justas; provides monitoring info) -> Unified (Jean-Roch; digests monitoring info into a per-task policy) -> JobRouter (Brian; converts per-task policy into set of rules for altering jobs)

> > Given there is no equivalent of Unified here, I would suggest:
> >
> > gwmsmon (Justas; provides monitoring info per task) -> JobRouter (Todor? converts monitoring to JobRouter rules)
> >
> > > > Put in a small PR on CRABServer to differentiate between wall time for matchmaking and wall time for killing jobs. That way, we can be more accurate without potentially affecting users.
> > >
> > > I saw it, thanks. That's really helpful.
> >
> > Very small first step!
> >
> > (Although, admittedly not a particularly long road!)
> >
> > Brian

— Re: Auto-tuning user jobs.eml

Subject: RE: Auto-tuning user jobs
From: Todor Trendafilov Ivanov <todor.trendafilov.ivanov@cern.ch>
Date: 19/04/17 10:46
To: Stefano Belforte <Stefano.Belforte@cern.ch>, "bbockelm@cse.unl.edu" <bbockelm@cse.unl.edu>
CC: Jean-Roch Vlimant <jean-roch.vlimant@cern.ch>

Hi Brian, Stefano, Jean,

I am already reading the documentation. Diego gave me the link with the basics from the htcondor manual. Give me a day to get in touch with the problem and then I will try to set a testing configuration in qa. I am sure that there are going to be alot of questions about the "information flow":

> > gwmsmon (Justas; provides monitoring info per task) -> JobRouter (Todor? converts monitoring to JobRouter rules)

or I should have at least an example of a task converting rule that Brain is setting after Unified - I think this is going to be helpful.

Regards,
Todor

____
From: Stefano Belforte
Sent: 19 April 2017 07:35
To: bbockelm@cse.unl.edu
Cc: Stefano Belforte; Jean-Roch Vlimant; Todor Trendafilov Ivanov
Subject: Re: Auto-tuning user jobs

Todor should be able to start working on this starting this
week or next one at latest. Hopefully he and Diego Davila
can start by understanding how current JobRouter-for-LPC works.
I am a bit lost at where we are on the gwmsmon part,
although it came to my mined that we can ask Emilis
to provide the up-to-date timing
info as a file on the schedd spool director via the

TaskProcess. That TP npw runs every 5 min and parses
condor log for the task to update on jobs status.
That would make the whole thing local to the schedd,
while the gwmsmon approach has more generality.

Stefano

On 19/04/17 05:46, Brian Bockelman wrote:

> Hi,
>
> I'd like to proceed along this route. How can we best move things forward?
>
> Thanks,
>
> Brian
>
> Sent from my iPhone
>
>> On Apr 10, 2017, at 10:00 AM, Brian Bockelman <bbockelm@cse.unl.edu> wrote:
>>
>>> On Apr 10, 2017, at 2:29 AM, Stefano Belforte <stefano.belforte@cern.ch> wrote:
>>>
>>>> On 04/09/2017 04:56 PM, Brian Bockelman wrote:
>>>> This would be done with the jobrouter - we do the same for the production systems.
>>>> Jobrouter is known to scale well for this.
>>>
>>> we have a JobRrouter on each CRAB schedd's to allow submission to LPC.
>>> https://gitlab.cern.ch/ai/it-puppet-hostgroup-vocmsglidein/blob/master/code/templates/modules/condor/90_cmslpc_jobrouter.config
>>> I know nothing about its configuration, usage etc. Should we get SI to deal
>>> with this ? Crab Cat-A operator, Todor, has to be on top eventually,
>>> but needs traning. Who is at CERN who can guide him on this ?
>>
>> Likely some combination of Jean-Roch and Justas.
>>
>> I can help if someone could give my CERN account (bbockelm) permission to view this repository!
>>
>>> Is JeanRoch in the thread so he can provide the code used in production
>>> as a start ?
>>
>> Yup! Basically, the information flow is:
>>
>> gwmsmon (Justas; provides monitoring info) -> Unified (Jean-Roch; digests monitoring info into a per-task policy) -> JobRouter (Brian; converts per-task policy into set of rules for altering jobs)
>>
>> Given there is no equivalent of Unified here, I would suggest:
>>
>> gwmsmon (Justas; provides monitoring info per task) -> JobRouter (Todor? converts monitoring to JobRouter rules)
>>
>>>> Put in a small PR on CRABServer to differentiate between wall time for matchmaking and wall time for killing jobs. That way, we can be more accurate without potentially affecting users.
>>>
>>> I saw it, thanks. That's really helpful.

> Very small first step!
>
> (Although, admittedly not a particularly long road!)
>
> Brian

— Re: Auto-tuning user jobs.eml

Subject: Re: Auto-tuning user jobs
From: Stefano Belforte <stefano.belforte@cern.ch>
Date: 19/04/17 12:45
To: Todor Trendafilov Ivanov <todor.trendafilov.ivanov@cern.ch>, "bbockelm@cse.unl.edu" <bbockelm@cse.unl.edu>
CC: Jean-Roch Vlimant <jean-roch.vlimant@cern.ch>, Justas Balcas <justas.balcas@cern.ch>

Brian,
to fix ideas. I learn from the other thread that Justas now has what looks the needed API, e.g. given CRAB task with these ads:

```
CRAB_UserHN CRAB_ReqName
smitra      170411_132805:smitra_crab_DYJets
```

One woudl lookup:
https://cms-gwmsmon.cern.ch/analysisview/json/historynew/percentileruntime720/smitra/170411_132805%3Asmitra_crab_DYJets

which return a json file

```
{"hits": {"hits": [], "total": 263, "max_score": 0.0}, "_shards": {"successful": 31, "failed": 0, "total": 31}, "took": 40, "aggregations": {"2": {"values": {"5.0": 6.543611111111126, "25.0": 11.444305555555555, "1.0": 3.5115222222222222, "95.0": 19.811305555555556, "75.0": 16.773194444444446, "99.0": 20.513038888888889, "50.0": 13.365277777777777}}}, "timed_out": false}
```

in hopefully fixed forever format so that the "values" can be extracted and one would e.g. pick the 95.0 one (i.e. 19.8 hours) multiply by 2, cap at 46h and set it.

Am I wrong somewhere ?

By the way, I'd add the additional requirement that we do this only on tasks which present the detault wall time requirement of 24h. When user bothered to indicate a different value, I'd respect it, if nothing else, to allow resubmitting that one job on the obnoxious horribly long long lumi which requres 12h when everything else is 1h, w/o CRAB resetting it to 2h over and over !

Differently from Production, in Analysis we always need to leave a door open for the users to take care themselves, it is the best way for edge cases.

Stefano

— Re: Auto-tuning user jobs.eml

Subject: Re: Auto-tuning user jobs
From: Brian Bockelman <bbockelm@cse.unl.edu>
Date: 19/04/17 20:56

To: Stefano Belforte <stefano.belforte@cern.ch>
CC: Todor Trendafilov Ivanov <todor.trendafilov.ivanov@cern.ch>, "Jean-Roch Vlimant" <jean-roch.vlimant@cern.ch>, Justas Balcas <justas.balcas@cern.ch>

> On Apr 19, 2017, at 5:45 AM, Stefano Belforte <stefano.belforte@cern.ch> wrote:
>
> Brian,
> to fix ideas. I learn from the other thread that Justas now has what
> looks the needed API, e.g. given CRAB task with these ads:
> CRAB_UserHN CRAB_ReqName
> smitra   170411_132805:smitra_crab_DYJets
>
> One woudl lookup:
> https://cms-gwmsmon.cern.ch/analysisview/json/historynew/percentileruntime720/smitra/170411_132805%3Asmitra_crab_DYJets
>
> which return a json file
> {"hits": {"hits": [], "total": 263, "max_score": 0.0}, "_shards": {"successful": 31, "failed": 0, "total": 31}, "took": 40, "aggregations": {"2": {"values": {"5.0": 6.5436111111111126, "25.0": 11.444305555555555, "1.0": 3.5115222222222222, "95.0": 19.811305555555556, "75.0": 16.773194444444446, "99.0": 20.513038888888889, "50.0": 13.365277777777777}}}, "timed_out": false}
>
> in hopefully fixed forever format so that the "values" can be extracted
> and one would e.g. pick the 95.0 one (i.e. 19.8 hours) multiply by 2, cap at
> 46h and set it.
>
> Am I wrong somewhere ?

I'd like a bit more elaborate (especially as we have worked to differentiate the "estimated runtime" from the "time at which we kill the job"):
- If less than 20 jobs have finished, do nothing!
- If at least 20 jobs have finished, take the 95th percentile and set estimated run time as min(95th percentile, user-provided runtime).

> By the way, I'd add the additional requirement that we do this only on
> tasks
> which present the detault wall time requirement of 24h. When user
> bothered to
> indicate a different value, I'd respect it, if nothing else, to allow
> resubmitting that one job on the obnoxious horribly long long lumi which
> requres
> 12h when everything else is 1h, w/o CRAB resetting it to 2h over and over !

> Differently from Production, in Analysis we always need to leave a door open
> for the users to take care themselves, it is the best way for edge cases.

But I think the important thing is to allow the user to specify when the job should be *killed* versus the estimate for scheduling purposes.

Here's the equivalent code used in production:

https://github.com/CMSCompOps/WmAgentScripts/blob/master/Unified/go_condor.py#L168

There, Unified pre-computes the final number - "go_condor.py" is simply converting the Unified data to a JobRouter configuration. Here, we would be doing the calculation in the script based on the raw data from gwmsmon.

Todor - is this making sense enough to start the work? Do you feel like you understand the existing JobRouter code enough to put in a PR? Any other questions?

Brian

— Re: Auto-tuning user jobs.eml

Subject: Re: Auto-tuning user jobs
From: Stefano Belforte <stefano.belforte@cern.ch>
Date: 20/04/17 01:00
To: Brian Bockelman <bbockelm@cse.unl.edu>
CC: <stefano.belforte@cern.ch>, Todor Trendafilov Ivanov <todor.trendafilov.ivanov@cern.ch>, Jean-Roch Vlimant <jean-roch.vlimant@cern.ch>, Justas Balcas <justas.balcas@cern.ch>

Yes. I am slowly getting things, thanks.
One question below:

On 19/04/17 20:56, Brian Bockelman wrote:

> I'd like a bit more elaborate (especially as we have worked to differentiate the "estimated runtime" from the "time at which we kill the job"):
> - If less than 20 jobs have finished, do nothing!
> - If at least 20 jobs have finished, take the 95th percentile and set estimated run time as min(95th percentile, user-provided runtime).

and I guess the number of jobs is the "total" in the json ?
i.e. in the example [1] is 263.
And I further guess that I do not need to know what _shards are
nor all the other things there and timed_out is likely referred
to the ES query, not the jobs.
So the code should avoid processing things with timed_out=true ?

thanks
Stefano

[1]

{"hits": {"hits": [], "total": 263, "max_score": 0.0}, "_shards": {"successful": 31, "failed": 0, "total": 31}, "took": 40, "aggregations": {"2": {"values": {"5.0": 6.5436111111111126, "25.0": 11.444305555555555, "1.0": 3.5115222222222222, "95.0": 19.811305555555556, "75.0": 16.773194444444446, "99.0": 20.513038888888889, "50.0": 13.365277777777777}}}, "timed_out": false}

which came from

https://cms-gwmsmon.cern.ch/analysisview/json/historynew/percentileruntime720/smitra/170411_132805%3Asmitra_crab_DYJets

— Re: Auto-tuning user jobs.eml

Subject: Re: Auto-tuning user jobs
From: Stefano Belforte <stefano.belforte@cern.ch>
Date: 20/04/17 01:10
To: Brian Bockelman <bbockelm@cse.unl.edu>
CC: <stefano.belforte@cern.ch>, Todor Trendafilov Ivanov <todor.trendafilov.ivanov@cern.ch>, Jean-Roch Vlimant <jean-roch.vlimant@cern.ch>, Justas Balcas <justas.balcas@cern.ch>

I think it is time to move this thread to some kind
of persistent place. So I have started this

https://github.com/dmwm/CRABServer/wiki/AUTO-TUNING-of-jobs-time-limit

did not have time to put in full info yet.

Will surely appreciate being pointing out what I have
already misunderstood about "the plan".

I'd like further discussion of this to happen in a forum,
e.g. hn-cms-cradevelopment.

Stefano

On 19/04/17 20:56, Brian Bockelman wrote:

> > On Apr 19, 2017, at 5:45 AM, Stefano Belforte <stefano.belforte@cern.ch <mailto:stefano.belforte@cern.ch>> wrote:
> >
> > Brian,
> > to fix ideas. I learn from the other thread that Justas now has what
> > looks the needed API, e.g. given CRAB task with these ads:
> > CRAB_UserHN CRAB_ReqName
> > smitra   170411_132805:smitra_crab_DYJets
> >
> > One woudl lookup:
> > https://cms-gwmsmon.cern.ch/analysisview/json/historynew/percentileruntime720/smitra/170411_132805%3Asmitra_crab_DYJets
> >
> > which return a json file
> > {"hits": {"hits": [], "total": 263, "max_score": 0.0}, "_shards": {"successful": 31, "failed": 0, "total": 31}, "took": 40, "aggregations": {"2": {"values": {"5.0": 6.543

> > 6111111111126, "25.0": 11.444305555555555, "1.0": 3.5115222222222222, "95.0": 19.811305555555556, "75.0": 16.773194444444446, "99.0": 20.513038888888889, "50.0": 13.365277777777777}}}, "timed_out": false}
> >
> > in hopefully fixed forever format so that the "values" can be extracted
> > and one would e.g. pick the 95.0 one (i.e. 19.8 hours) multiply by 2, cap at 46h and
> > set it.
> >
> > Am I wrong somewhere ?
>
> I'd like a bit more elaborate (especially as we have worked to differentiate the "estimated runtime" from the "time at which we kill the job"):
> - If less than 20 jobs have finished, do nothing!
> - If at least 20 jobs have finished, take the 95th percentile and set estimated run time as min(95th percentile, user-provided runtime).
>
> > By the way, I'd add the additional requirement that we do this only on tasks
> > which present the detault wall time requirement of 24h. When user bothered to
> > indicate a different value, I'd respect it, if nothing else, to allow
> > resubmitting that one job on the obnoxious horribly long long lumi which requres
> > 12h when everything else is 1h, w/o CRAB resetting it to 2h over and over !
> >
> > Differently from Production, in Analysis we always need to leave a door open
> > for the users to take care themselves, it is the best way for edge cases.
>
> But I think the important thing is to allow the user to specify when the job should be *killed* versus the estimate for scheduling purposes.
>
> Here's the equivalent code used in production:
>
> https://github.com/CMSCompOps/WmAgentScripts/blob/master/Unified/go_condor.py#L168
>
> There, Unified pre-computes the final number - "go_condor.py" is simply converting the Unified data to a JobRouter configuration. Here, we would be doing the calculation in the script based on the raw data from gwmsmon.
>
> Todor - is this making sense enough to start the work? Do you feel like you understand the existing JobRouter code enough to put in a PR? Any other questions?
>
> Brian

Attachments:

Auto-tuning user jobs.eml	10.1 KB
Re: Auto-tuning user jobs.eml	1.8 KB
Re: Auto-tuning user jobs.eml	13.2 KB
Re: Auto-tuning user jobs.eml	2.9 KB
Re: Auto-tuning user jobs.eml	3.8 KB
Re: Auto-tuning user jobs.eml	12.3 KB
Re: Auto-tuning user jobs.eml	3.6 KB
Re: Auto-tuning user jobs.eml	4.7 KB
Re: Auto-tuning user jobs.eml	2.0 KB

Re: Auto-tuning user jobs.eml	4.3 KB
Re: Auto-tuning user jobs.eml	3.2 KB
Re: Auto-tuning user jobs.eml	14.4 KB
Re: Auto-tuning user jobs.eml	4.0 KB
Re: Auto-tuning user jobs.eml	4.8 KB
Re: Auto-tuning user jobs.eml	3.0 KB
Re: Auto-tuning user jobs.eml	18.4 KB
Re: Auto-tuning user jobs.eml	2.9 KB
Re: Auto-tuning user jobs.eml	4.7 KB