Open Science Grid

How to get the needed computing

Tuesday afternoon, 1:30pm

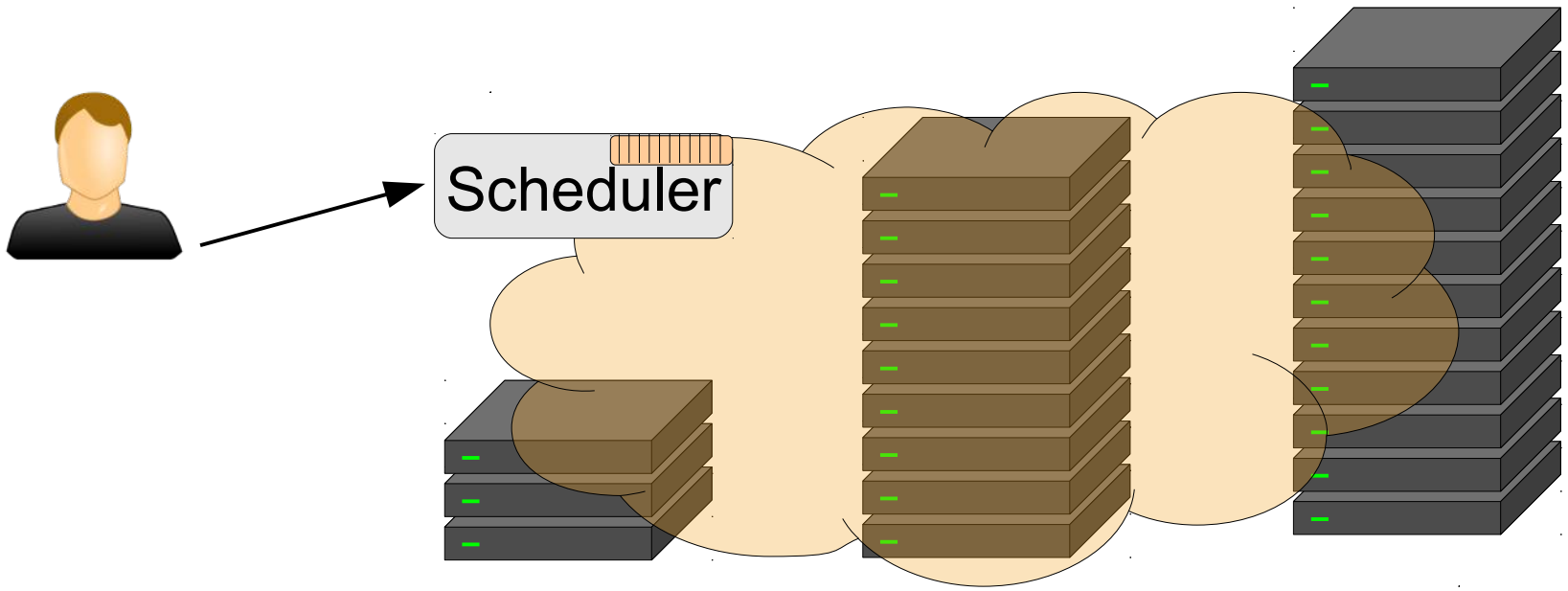
Igor Sfiligoi <isfiligoi@ucsd.edu>
Leader of the OSG Glidein Factory Operations
University of California San Diego

Logistical reminder

- It is OK to ask questions
  - During the lecture
  - During the demos
  - During the exercises
  - During the breaks
- If I don't know the answer,
  I will find someone who likely does

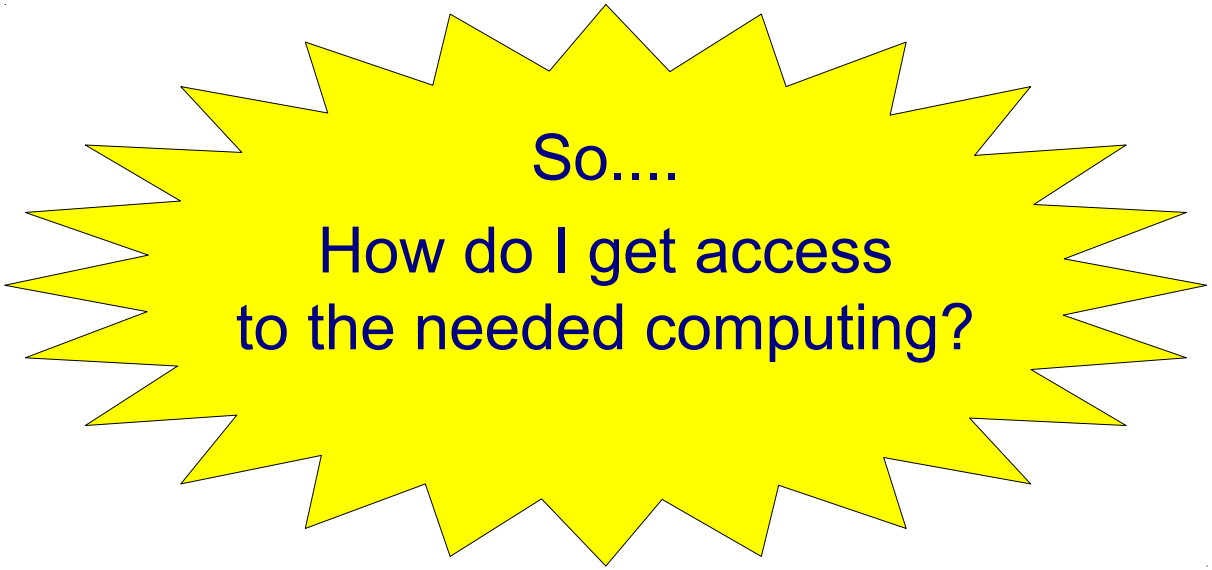
Reminder - (D)HTC

- (D)HTC is about getting as many CPU cycles as possible over the long run
  - No matter where they are located



But where do you get the CPUs?

- You likely have a desktop/laptop
- But that's normally not enough
  - Or you wouldn't be here!



Getting the needed CPUs

You either:

- Buy it
- Get a computing grant
- Use leftover CPU cycles

Getting the needed CPUs

You either:

- Buy it
  - Your own hardware (e.g. a cluster)
  - Contribute hardware to a common pool
  - Lease machines on monthly/yearly basis
  - Rent machines on a per hour basis
- Get a computing grant
- Use leftover CPU cycles
  - On friends' hardware
  - At your home institution
  - On a large-scale scientific infrastructure

Getting the needed CPUs

You either:

- Buy it
  - Your own hardware (e.g. a cluster)
  - Contribute hardware to a common pool
  - Lease machines on monthly/yearly basis
  - Rent machines on a per hour basis
- Get a computing grant
- Use leftover CPU cycles
  - On friends' hardware
  - At your home institution
  - On a large-scale scientific infrastructure

Which way to go?

- Buying is more reliable
  - You can plan on how much computing you will be able to do
  - But not always an option
    (We all have a limited budget)
- Grants are similar in nature
  - But may be harder to get than money!
- Opportunistic use can give you vastly more
  - But you will have to be flexible
  - And there is no guarantee you will get what you hope for

Buying the hardware

- Buying your own hardware is the most straightforward approach
- But you must also budget for
  - Floor space
  - Electricity
  - Cooling
  - Personnel costs

  } Often more than the HW itself
- You will need significant expertise

Buying the hardware

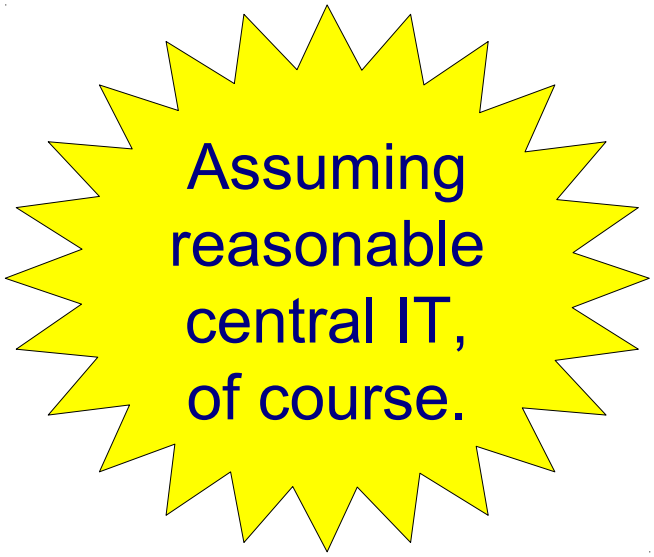
- Buying your own hardware is the most straightfo
- B
  - itself
  -
- You

And unless you have very constant compute needs, you must budget for peak use.

What will happen to CPUs when you don't use them?

Contributing to a common pool

- If you have an existing cluster at your institution/campus, it **may be cheaper** to contribute to it
  - Economies of scale
  - Better expertise
- You will have to trade some flexibility
  - But often it is well worth it



Leveling off the spikes

- Most people have spiky compute needs
  - I have a great idea...
    need a gazillion CPUs **now** to verify it!
    Then nothing for a month or more.
  - Everybody wants to run a last computing pass just before that important conference
- Planning for the average and leasing the rest during peak demand may be more efficient

Handling the spikes

- Again, you can either buy or use opportunistic resources
  - Same considerations as before
- If you are willing to pay
  - Cloud computing provides quick access on a pay-as-you-go basis
  - You can get lower rates if you lease machines on monthly basis

Cloud computing

- You pay by the hour
  - And are given a dedicated (virtual) machine all for yourself
- Most famous is Amazon EC2
  (but not the only one)
- There are also companies that will create a HTC system on top of them
  - e.g. Cycle Computing's CycleCloud

Cloud computing

- You pay by t[illegible] h[illegible]
  - A[illegible]

You may hear talking about “Scientific Clouds”.

That's really just another name for Grid+VMs

(so not what I call Cloud here)

Server hosting

- You can also lease nodes on a longer term basis
  - Typically 3+ months
  - Popular vendor: Rackspace
- Similar to the Cloud approach
  - But requires more planning
  - And is typically cheaper
    - Assuming you use the CPUs 24x7

Grants

- Some of the US large-scale computing infrastructures are grant based
  - e.g. XSEDE
- You make a proposal, and if you make a good case, you can get a substantial CPU allocation
  - Be prepared for a lot of paperwork
  - And long lead times

Opportunistic use

- If you need more CPU than you can afford to buy
  - Consider using opportunistic CPU cycles
  - i.e. available CPU not needed by the owners
- Remember, there is no guarantee you will get anything that way
  - But still very likely you will get a lot!

Opportunistic resources

- You should first look close by
  - Cluster(s) at your home institution may have lots of idle CPU
  - You may not even need to change a thing!
- Once that is not enough, large scale compute infrastructures can provide lots of CPU cycles
  - e.g. OSG

Beggars can't be choosers

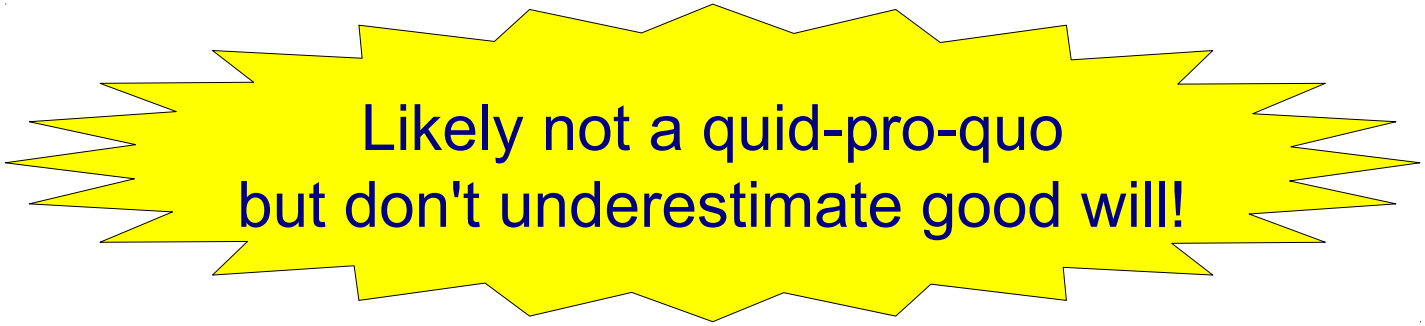
- Remember, when you go the opportunistic route, you have very few rights
  - You are effectively “a beggar”
- You will have to adapt
  - Don't expect to get it your way
    - Even though some sites may be willing to help
  - The more flexible you are, the more sites you will be able to use

The benefit of Overlays

- You will likely end up with a mix of owned, leased and borrowed resources
- Using an overlay system makes life easier
  - Both for job scheduling, and
  - Leveling any differences in the resources

Consider contributing back

- If you buy hardware, consider contributing back
  - i.e. slightly overprovision your cluster and give unused CPU cycles to others
- A modest amount of gifted CPU can buy you a lot of good will
  - For when you need to level your own spikes



Questions?

- Questions? Comments?
  - Feel free to ask me questions later:

    Igor Sfiligoi <isfiligoi@ucsd.edu>
- Upcoming sessions
  - Now – 2:15pm
    - Demo
  - 2:15pm – 3:00pm
    - Next lecture - OSG & XSEDE – by Alain Roy
  - 3:00pm – 3:15pm
    - Break

Grids and Clouds