# How to get the needed computing

## Tuesday morning, 11:15am

Igor Sfiligoi <isfiligoi@ucsd.edu>

Leader of the OSG Glidein Factory Operations

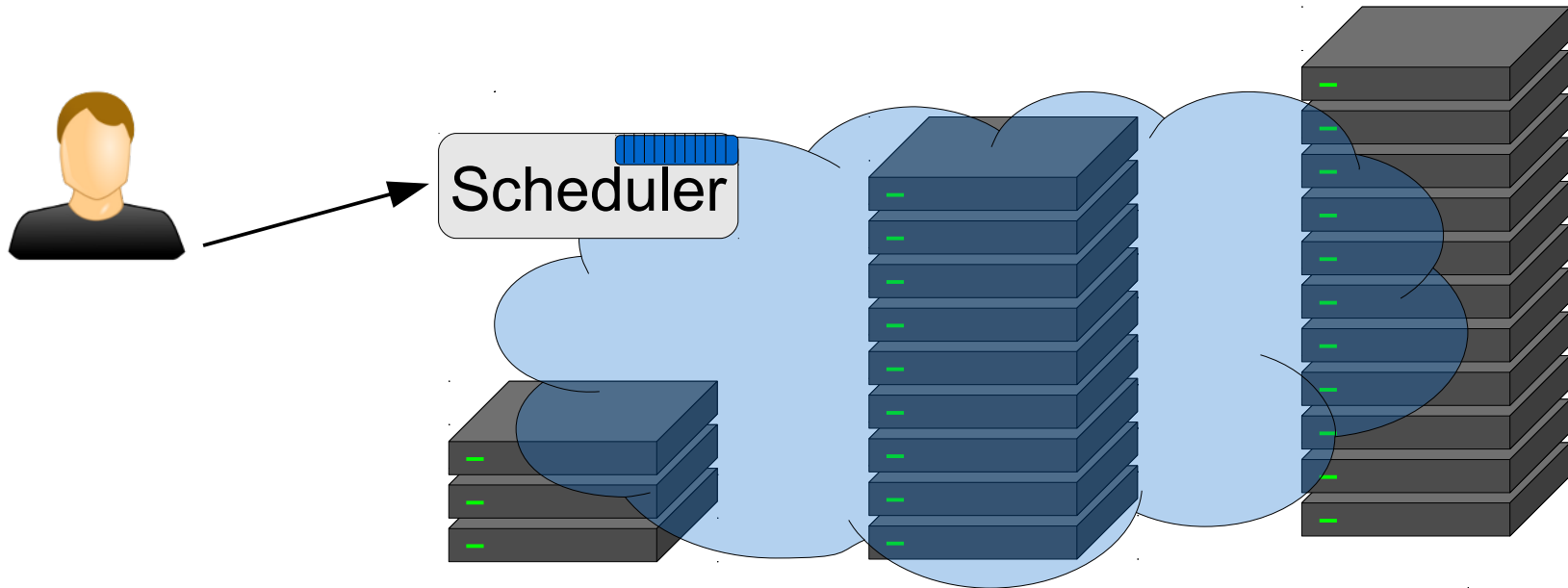University of California San Diego

# 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!

So....

How do **I** get access
to the needed computing?

# Getting the needed CPUs

## You either:

- Buy it

- Get a computing grant
- Use leftover CPU cycles

Getting the computing

# 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

Getting the computing

# 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

Getting the computing

# Don't assume flat usage

- While HTC paradigm postulates maximizing results out of available resources

- Most people really have spiky compute needs, e.g.

  - You come up with a great idea...
    need a gazillion CPUs **now** to verify it!
    Then nothing for a month or more
    while you look for the next great one.

  - Everybody wants to run a last computing pass just before that important conference

# Impact on planning

- You likely cannot afford to buy enough dedicated resources to cover the peaks

- You should **always** plan on using leased resources
  - At least for part of the year
  - i.e. make sure you choose tools that allow you to do DHTC

# Buying dedicated hardware

- Buying your own hardware is the most straightforward approach

- But you must also budget for
  - Floor space
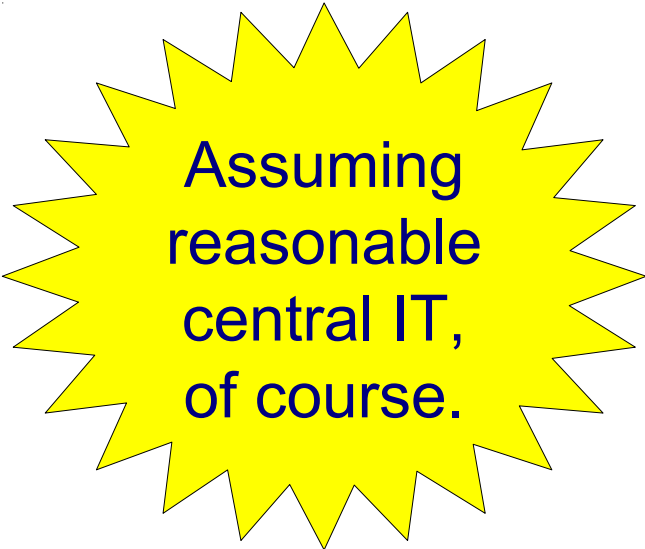  - Electricity
  - Cooling
  - Personnel costs

  Often more than the HW itself

- And install your own HTC system
  - Which requires significant expertise

Getting the computing

# Server hosting

- You can lease hardware from commercial entities
    - Typically 3-12 months leases
    - Popular vendor: Rackspace
- Great for shorter projects
    - Likely gives you lower Total Cost of Ownership
        - May cost less than buying the HW itself
        - And you save on the infrastructure costs
    - But still requires you to operate an HTC system

# Contributing to a common pool

- If you have an existing HTC system 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

Assuming reasonable central IT, of course.

# Contributing to a common pool

- If you h̶a̶... ...ystem at v̶... ...

The HTC system
may also be located
at a different location,
if they allow remote access.

Usually called
Grid computing

# Computing 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 on their HTC system

- Initial effort comparable to buying HW
  - A lot of paperwork needed
  - And long lead times

# Cloud computing

- Similar to server hosting

  - But you pay by the hour

- Most famous is Amazon EC2
  (but not the only one)

- Great for spike leveling

  - Can get a lot of resources on short notice, if you have the needed money

  - But can be quite expensive

# Cloud computing

- Similar to server hosting

You may hear talking about "Scientific Clouds".

That's really just another name for VM-based HTC systems

(so **not** what I call Cloud here)

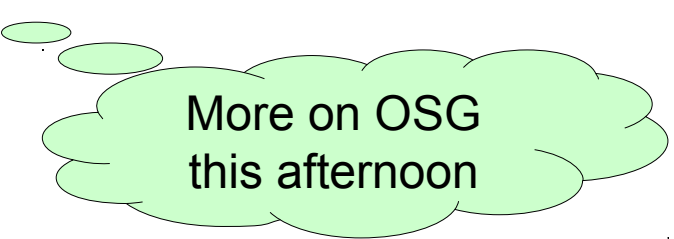Getting the computing

# Going for opportunistic resources

- All the methods described so far give you dedicated resources
  - i.e. you can count on them being there when you need them
    (at least, after the initial setup period)
  - But that may not be enough for your needs
    - And you cannot afford more

- Opportunistic resources may came to the rescue
  - Just remember there are **no guarantees** here

# Opportunistic resources

- The opportunistic resources are essentially machines that are **currently** not needed by the owner
  - e.g. owners are in the low part of a spike
- Reasons why they may allow you to use them (instead of turning them off)
  - Connections – e.g. Friend's desktop
  - Politics – e.g. Funding agency requirement
  - Money – e.g. Amazon's spot instances

# Free HTC resources

- You should first look close by
  - The HTC system at your home institution may frequently have spare capacity
  - They are likely happy to share
- Once that is not enough, remote HTC clusters may provide substantial additional free resources as well
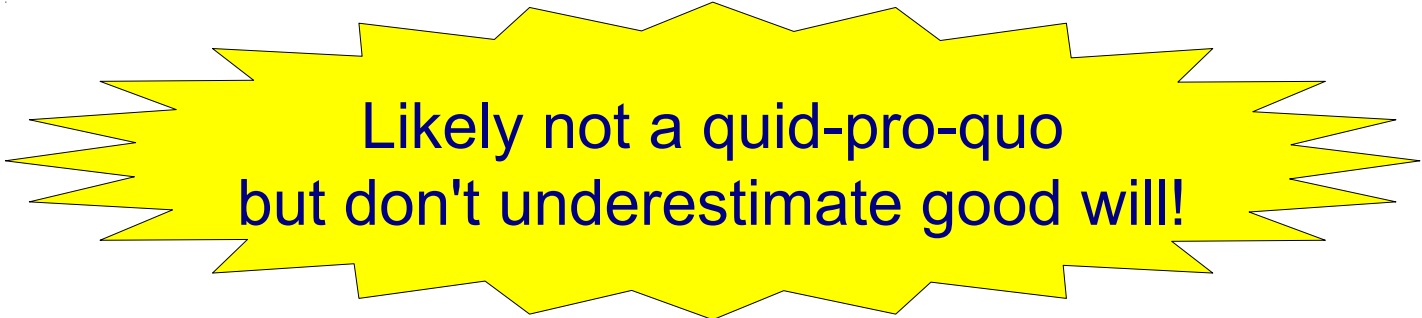  - e.g. Sites on OSG

More on OSG
this afternoon

Getting the computing

# Beggars can't be choosers

- Remember, when you scout for free resources, 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 sysadmins may be willing to help

    - The more flexible you are, the more free resources you will be able to use

Getting the computing

# 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

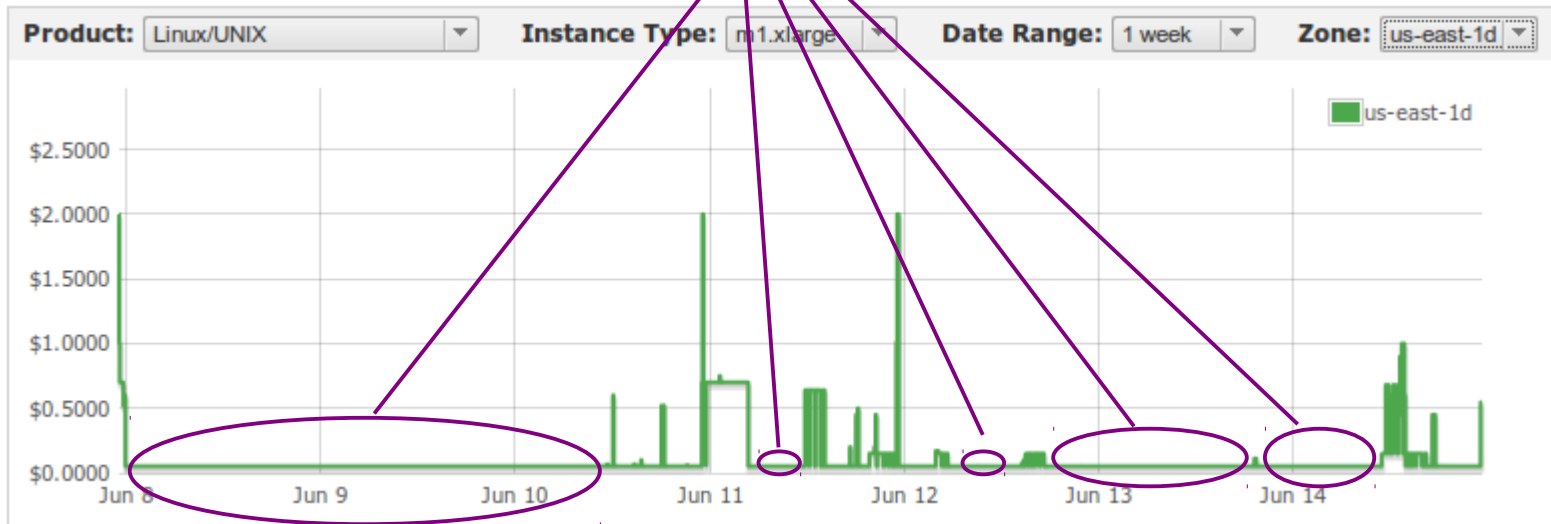  - For when you need to level your own spikes

Likely not a quid-pro-quo
but don't underestimate good will!

Getting the computing

# Opportunistic Cloud Resources

- Here opportunistic means **cheap(er)**
  - Not free

- Basically one bids (real money) for resources
  - The highest bidder gets the resources
  - Until someone else bids higher
  - e.g. Amazon EC2 Spot Instances

- Can get a lot for relatively little money
  - But not always

# Example

- ## For Amazon EC2 m1.xlarge
  - Full price: $0.5/hour
  - Min spot price: $0.05/hour
    - Many periods when you could get that price

Getting the computing

# Putting everything together

- You will get most work done if you mix owned, leased and opportunistic resources
  - i.e. DHTC

- Plan on using an overlay system early on
  - So you don't have to use N different tools to use N different clusters
  - Searching for/Learning about new tools when you are close to a deadline is no fun

# Questions?

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

    Igor Sfiligoi <isfiligoi@ucsd.edu>

- Upcoming sessions
  - Now – noon
    - Hands-on and demo
  - noon – 12:40pm
    - Tour
  - 12:40pm – 1:30pm
    - Lunch

# Grids and Clouds

Getting the computing