Profiling Applications to Choose the Right Computing Infrastructure plus Batch Management with HTCondor

Rob Quick <rquick@iu.edu>

OSG Operations Officer

Indiana University

Content Contributed by the University of Wisconsin Condor Team and Scot Kronenfeld

Open Science Grid

Follow Along at:

https://twiki.grid.iu.edu/bin/view/Education/RDASummerSchool/RDASchoolMaterials

Some thoughts on the exercises

- It's okay to move ahead on exercises if you have time
- It's okay to take longer on them if you need to
- If you move along quickly, try the "On Your Own" sections and "Challenges"

Open Science Grid

Most important!

- Please ask me questions!
  - …during the lectures
  - …during the exercises
  - …during the breaks
  - …during the meals
  - …over dinner
  - …via email after we depart (rquick@iu.edu)
- If I don't know, I'll find the right person to answer your question.

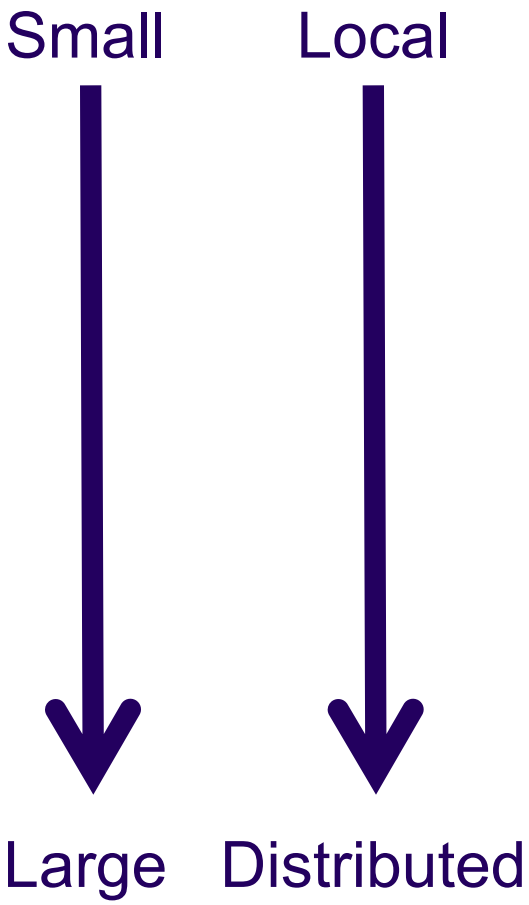
Quick UNIX Refresher Before We Start

- $, %, #
- ssh username@10.0.0.252
- nano, vi, emacs, cat >, etc.
- cd, mkdir, ls, rm, cp (scp)

Goals for this session

- Profiling your application
- Picking the appropriate resources
- Understand the basics of HTCondor

Let's take one step at a time



- Can you run one job on one computer?
- Can you run one job on another computer?
- Can you run 10 jobs on a set of computers?
- Can you run a multiple job workflow?
- How do we put this all together?

This is the path we'll take in the school

What does the user provide?

- A "headless job"
  - Not interactive/no GUI: how could you interact with 1000 simultaneous jobs?
- A set of input files
- A set of output files
- A set of parameters (command-line arguments)
- Requirements:
  - Ex: My job requires at least 2GB of RAM
  - Ex: My job requires Linux
- Control/Policy:
  - Ex: Send me email when the job is done
  - Ex: Job 2 is more important than Job 1
  - Ex: Kill my job if it runs for more than 6 hours

Open Science Grid

What does the system provide?

- Methods to:
  - Submit/Cancel job
  - Check on state of job
  - Check on state of available computers
- Processes to:
  - Reliably track set of submitted jobs
  - Reliably track set of available computers
  - Decide which job runs on which computer
  - Manage a single computer
  - Start up a single job

Open Science Grid

Gedankenexperiment

- Let's assume you have a 'large job'
  - What factors could make it large?
- Large Data Input or Output or both
- Needs a lot of calculation
- Needs a lot of memory
- Needs to communicate with other jobs (whether required or not)
- Reads and writes a lot of data
- Heavy graphics processing
- Any combination of any of the above

Open Science Grid

There is no "One Size Fits All Solution"

- But some solutions are more "Open" than others.
  - Local Laptop/Desktop
  - Local Cluster
  - HPC System
  - Shared HTC Resources
  - Clouds

Open Science Grid

Why is HTC hard?

- The HTC system has to keep track of:
  - Individual tasks (a.k.a. jobs) & their inputs
  - Computers that are available
- The system has to recover from failures
  - There will be failures! Distributed computers means more chances for failures.
- You have to share computers
    - Sharing can be within an organization, or between orgs
    - So you have to worry about security
    - And you have to worry about policies on how you share
- If you use a lot of computers, you have to handle variety:
  - Different kinds of computers (arch, OS, speed, etc..)
  - Different kinds of storage (access methodology, size, speed, etc…)
  - Different networks interacting (network problems are hard to debug!)

Open Science Grid

Surprise! Condor does this (and more)

- Methods to:
  - Submit/Cancel job. `condor_submit/condor_rm`
  - Check on state of job. `condor_q`
  - Check on state of avail. computers. `condor_status`
- Processes to:
  - Reliably track set of submitted jobs. `schedd`
  - Reliably track set of avail. computers. `collector`
  - Decide which job runs on where. `negotiator`
  - Manage a single computer `startd`
  - Start up a single job `starter`

Open Science Grid

But not only Condor

- You can use other systems:
  - PBS/Torque
  - Oracle Grid Engine (né Sun Grid Engine)
  - LSF
  - SLURM
  - …
- But I won't cover them.
  - Our expertise is with Condor
  - Our bias is with Condor
  - Overlays exist
- What should you learn at the school?
  - How do you think about HTC?
  - How can you do your science with HTC?
  - … For now, learn it with Condor, but you can apply it to other systems.

A brief introduction to Condor

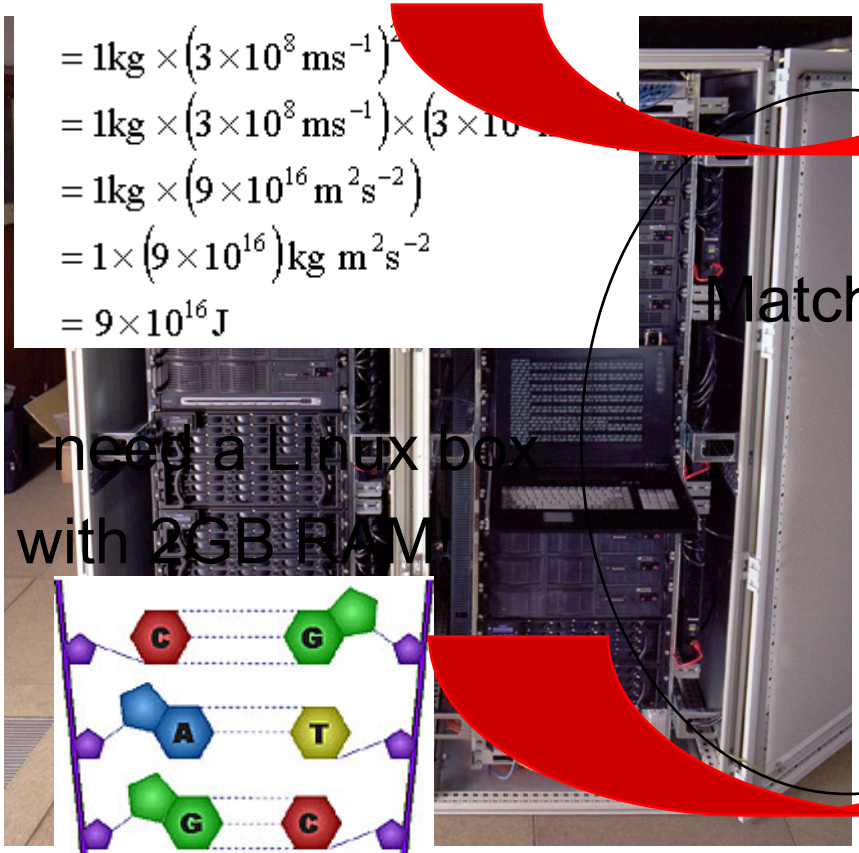
- Please note, we will only scratch the surface of Condor:
  - We won't cover MPI, Master-Worker, advanced policies, site administration, security mechanisms, submission to other batch systems, virtual machines, cron, high-availability, computing on demand, containers.

...And finds them Condor Takes Computers...

I need a Mac!

Desktop Computers

$$E = mc^2$$
$$= 1\text{kg} \times \left(3 \times 10^8 \text{ms}^{-1}\right)^2$$
$$= 1\text{kg} \times \left(3 \times 10^8 \text{ms}^{-1}\right) \times \left(3 \times 10^8 \text{ms}^{-1}\right)$$
$$= 1\text{kg} \times \left(9 \times 10^{16} \text{m}^2\text{s}^{-2}\right)$$
$$= 1 \times \left(9 \times 10^{16}\right) \text{kg m}^2\text{s}^{-2}$$
$$= 9 \times 10^{16} \text{J}$$



I need a Linux box with 2GB RAM!

Open Science Grid

Quick Terminology

- Cluster: A dedicated set of computers not for interactive use
- Pool: A collection of computers used by Condor
  - May be dedicated
  - May be interactive
- Remember:
  - Condor can manage a cluster in a machine room
  - Condor can use desktop computers
  - Condor can access remote computers
  - HTC uses all available resources

Open Science Grid

Matchmaking

- Matchmaking is fundamental to Condor
- Matchmaking is two-way
  - Job describes what it requires:

    I need Linux && 8 GB of RAM
  - Machine describes what it requires:

    I will only run jobs from the Physics department
- Matchmaking allows preferences
  - I need Linux, and I prefer machines with more memory but will run on any machine you provide me

Open Science Grid

Why Two-way Matching?

- Condor conceptually divides people into three groups:
  - Job submitters
  - Computer owners
  - Pool (cluster) administrator

} May or may not be the same people

- All three of these groups have preferences

Open Science Grid

ClassAds

- ClassAds state facts
  - My job's executable is analysis.exe
  - My machine's load average is 5.6
- ClassAds state preferences
  - I require a computer with Linux
- ClassAds are extensible
  - They say whatever you want them to say

Example ClassAd

```
MyType        = "Job"          ←— String
TargetType    = "Machine"
ClusterId     = 1377           ←— Number
Owner         = "roy“
Cmd           = “analysis.exe“
Requirements =                 ←— Expression
    (Arch == "INTEL")
&& (OpSys == "LINUX")
&& (Disk >= DiskUsage)
&& ((Memory * 1024)>=ImageSize)
…
```

Open Science Grid

Schema-free ClassAds

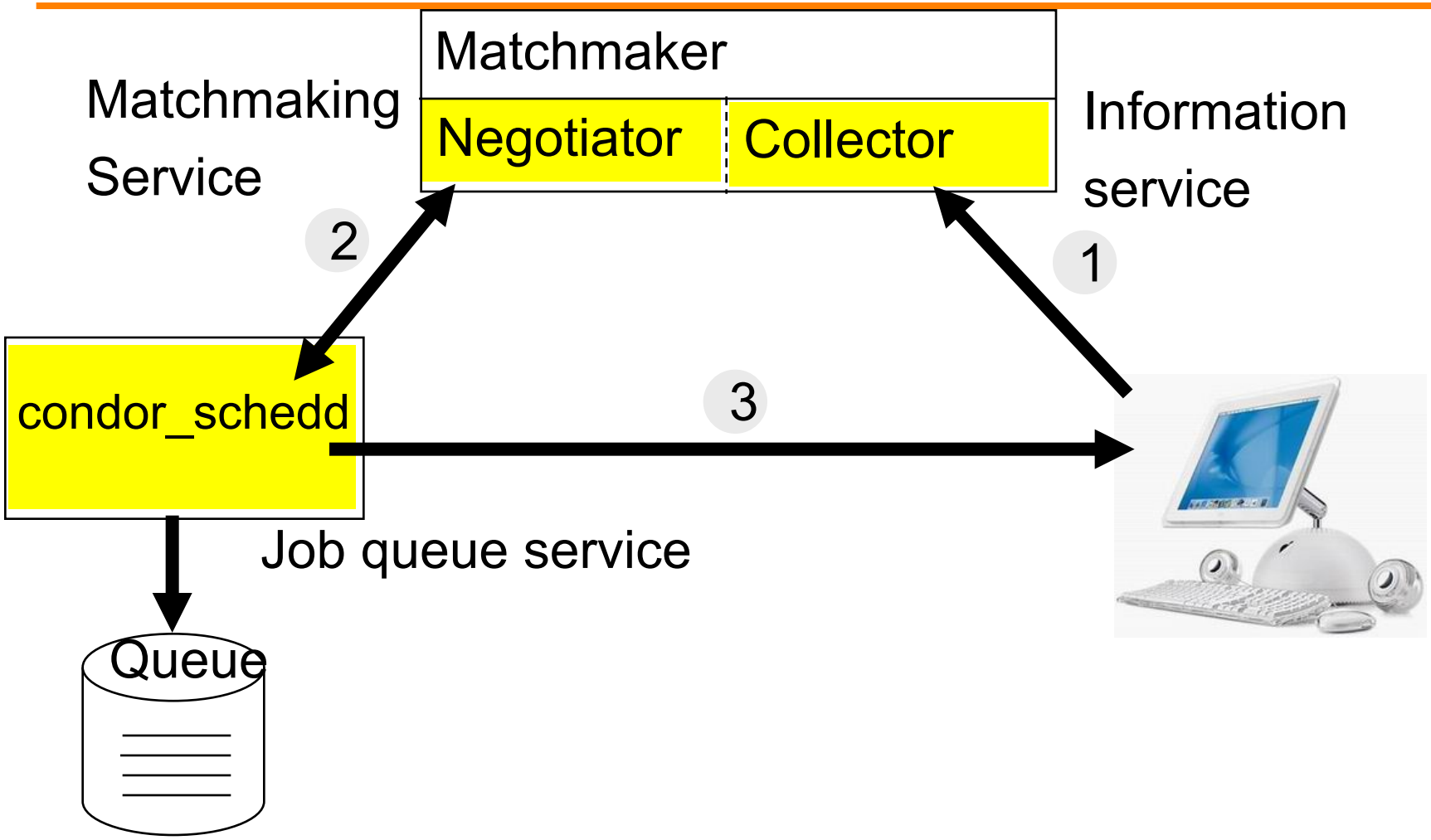
- Condor imposes some schema
  - Owner is a string, ClusterID is a number…
- But users can extend it however they like, for jobs or machines
  - `AnalysisJobType = "simulation"`
  - `HasJava_1_6 = TRUE`
  - `ShoeLength = 10`
- Matchmaking can use these attributes
  - `Requirements = OpSys == "LINUX"`
    `&& HasJava_1_6 == TRUE`

Open Science Grid

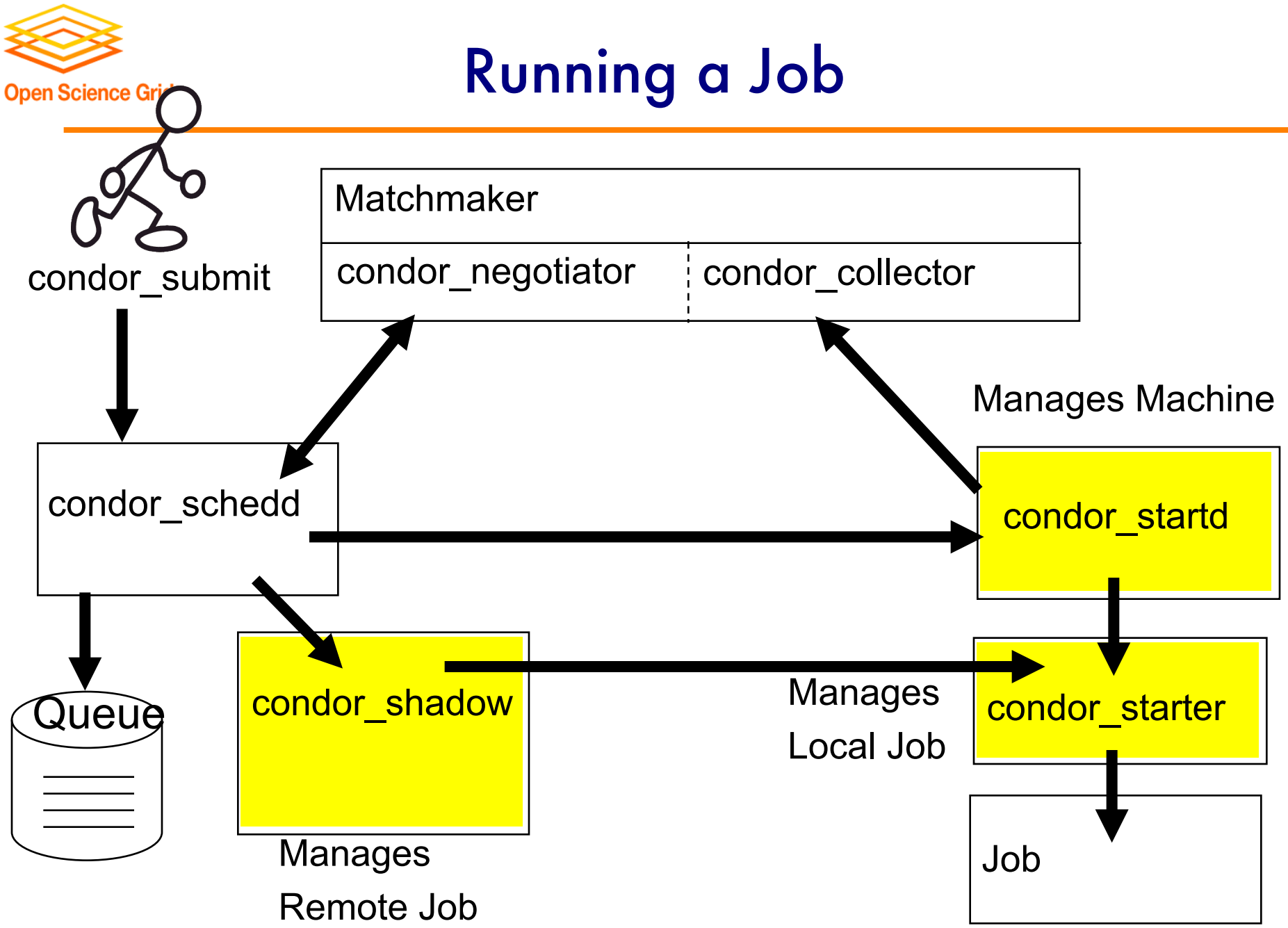
Don't worry

- You won't write ClassAds (usually)
  - You'll create a simple *submit file*
  - Condor will write the ClassAd
  - You can extend the ClassAd if you want to
- You won't write requirements (usually)
  - Condor writes them for you
  - You can extend them
  - In some environments you provide attributes instead of requirements expressions

Matchmaking diagram



Running a Job



That was a whirlwind tour!

- Enough with the presentation: let's use Condor!
- Goal: Extend the diversity of our jobs and add some data to the mix.

Some Day One Wrap Up Notes

- There are a lot of different compute environments out there
- There is a very diverse set of computing jobs out there
- Matching jobs to resources is key to not wasting resources
- Not all of the available environments are open environments

Open Science Grid

Questions?

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

    Rob Quick rquick@iu.edu

Exercises start here:

https://twiki.grid.iu.edu/bin/view/Education/RDASummerSchool/RDASchoolMaterials

Presentations are also available from this URL.