



# Communicating with Users

about HTCondor and High Throughput Computing

Lauren Michael, Research Computing Facilitator  
HTCondor Week 2015

# CHTC Computing

Center for **H**igh **T**hroughput **C**omputing, est. 2006

## > Large-scale, campus-shared computing systems

- campus high-throughput (HTC) grid and high-performance (HPC) cluster resources
- **all standard services provided free-of-charge**
- hardware buy-in options for *priority* access
- automatic access to the Open Science Grid (OSG)
- [chtc.cs.wisc.edu](http://chtc.cs.wisc.edu)



# CHTC-Accessible Computing:



**CHTC**

# CHTC-Accessible Computing:



**“UW Grid”**

**CHTC**



# CHTC-Accessible Computing:



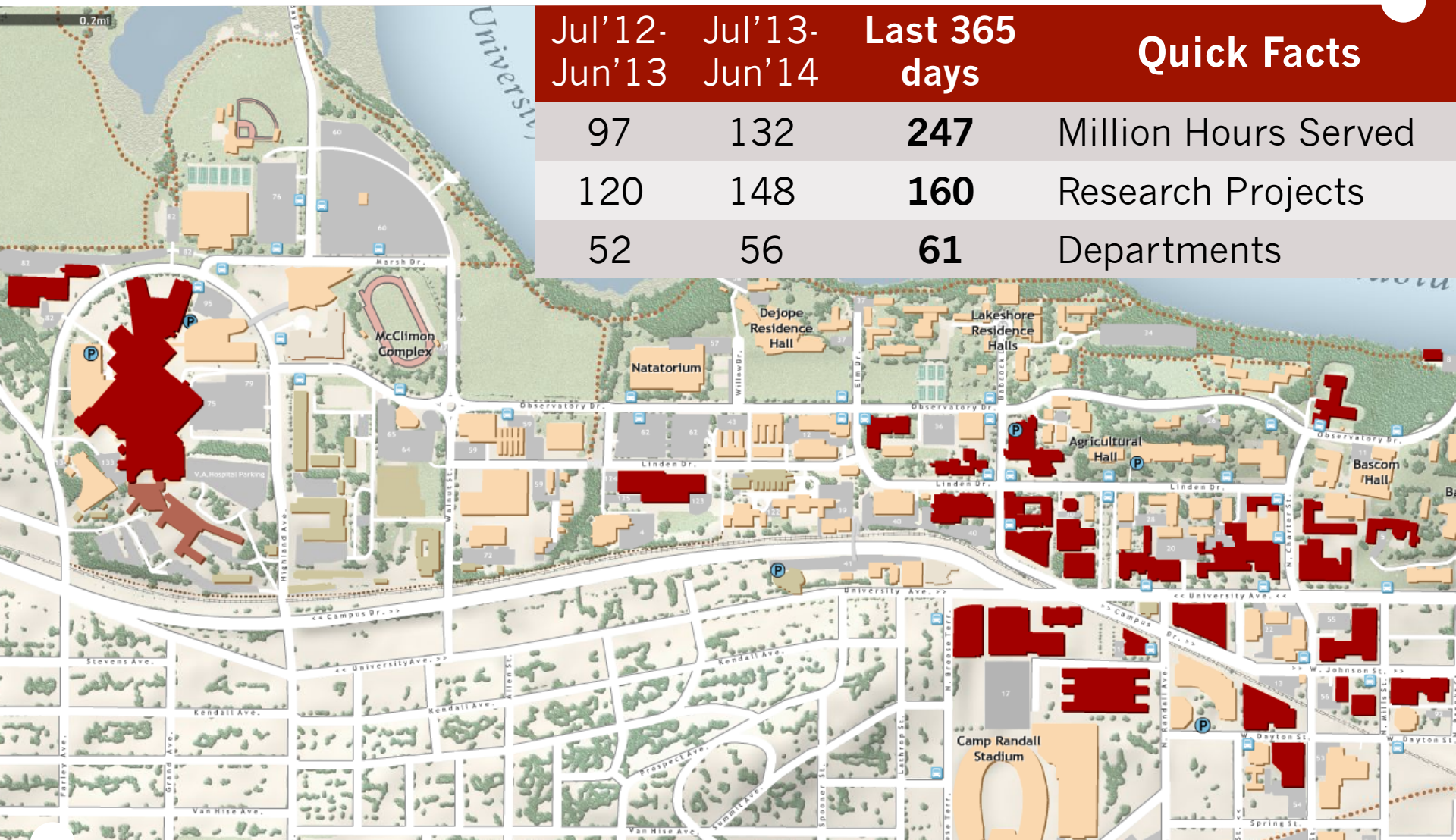
## Open Science Grid

[opensciencegrid.org](http://opensciencegrid.org)

“UW Grid”

CHTC





Researchers who use the CHTC are located all over campus (red buildings)



Jul'12- Jun'13	Jul'13- Jun'14	Last 365 days	Quick Facts
97	132	<b>247</b>	Million Hours Served
120	148	<b>160</b>	Research Projects
52	56	<b>61</b>	Departments

Individual researchers:  
**20+ years of computing**  
per day

Researchers who use the CHTC are located all over campus (red buildings)

# CHTC Services

- › **Support for using our compute systems**
  - consultation services, training, and proposal assistance
  - tools for numerous software (including Python, Matlab, R)
- › **Systems design/administration consulting**



# CHTC Services

- › **Support for using our compute systems**
  - consultation services, training, and proposal assistance
  - tools for numerous software (including Python, Matlab, R)
- › **Systems design/administration consulting**

## HTCondor: CHTC's R&D Arm

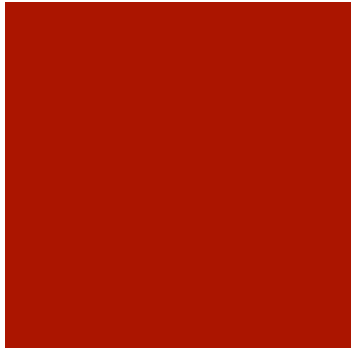
- › **Services provided to the campus community**
  - R&D for HTC Software
    - HTCondor, DAGMan (automated workflows), etc.
  - Software Engineering Expertise & Consulting
  - Software Testing & Security Consulting



# Problem:

Large-scale computing is complex, and  
not all users speak “computer geek”

Communicating *well* is hard.



**Users are people.**

# Know Your People



# Know Your People



## **1. What is the person's understanding of relevant terms?**

RAM, CPU, node, high-throughput computing (HTC)

## **2. What prior experience does the person come with?**

unix command line? programming? schedulers?

## **3. Is the person following what you say?**

Make it easy for researchers  
to find **the right people.**

## “Facilitators”

- consultants/liaisons for research computing
- identify with researchers
- identify with the *user* perspective



# Provide Clear Explanations



**Cater your communication to the person.**

(more difficult, but even more IMPORTANT, in email)

**Keep things simple.**

- Avoid unnecessary details, but allude to them.
- Start with the “big picture”

**Introduce new vocabulary when necessary.**

**Define terms. Be consistent.**

- Avoid “*terms of confusion*”

# Terms of Confusion:

## “high level”



# Terms of Confusion:

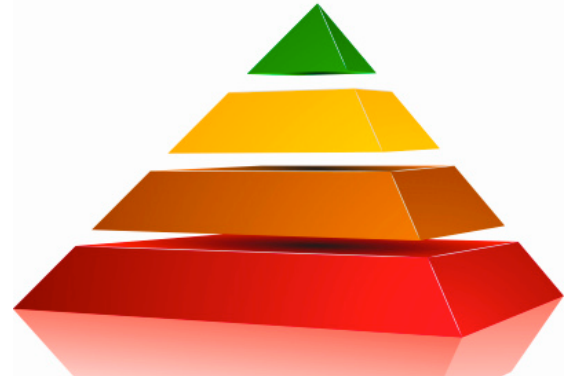
## “high level”

**Computer Geek:**



... of abstraction  
“bird’s eye view”  
“big picture”

**Many Users:**



... of complexity  
... of detail  
“advanced”



# Terms of Confusion:

## “high level” alternatives

“big picture”

“Basically, ...”

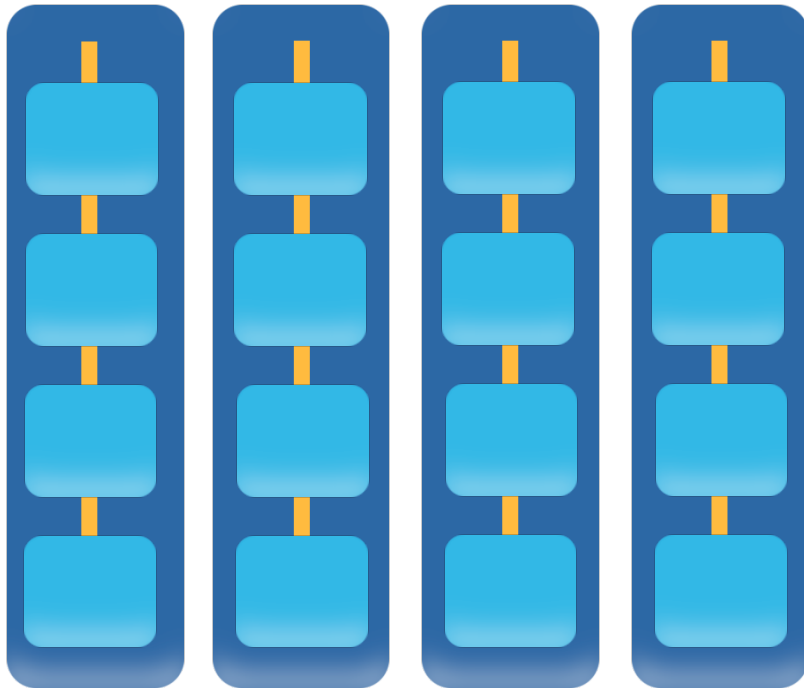
“If we step back ...”

or, just define what you mean

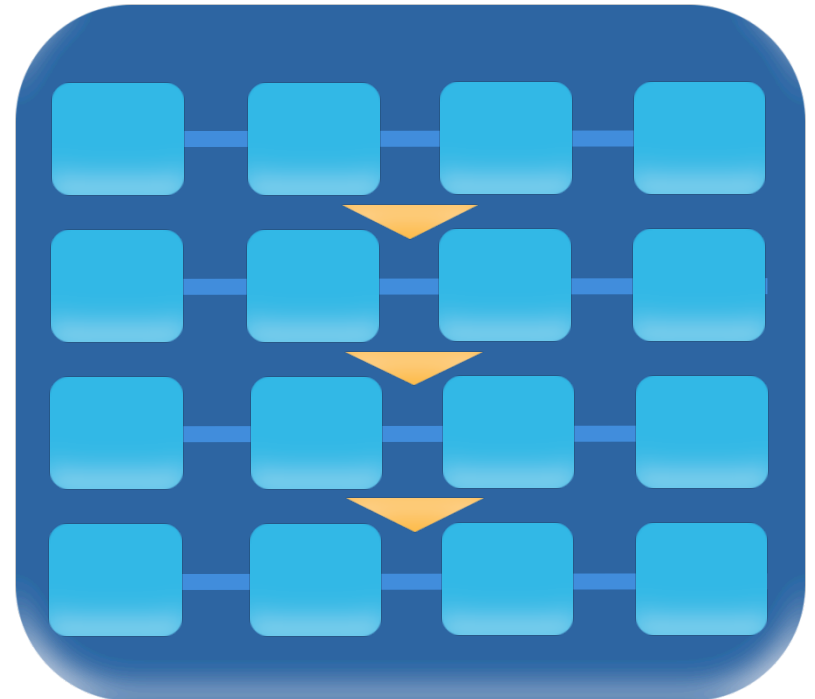
# Terms of Confusion:

“parallel”, “parallelize”

high-throughput



high-performance





# Terms of Confusion:

## “parallel” alternatives

“independent tasks”

“separate jobs”

**-versus-**

“parallelize within the program”

“multi-thread”, “MP”, “MPI”

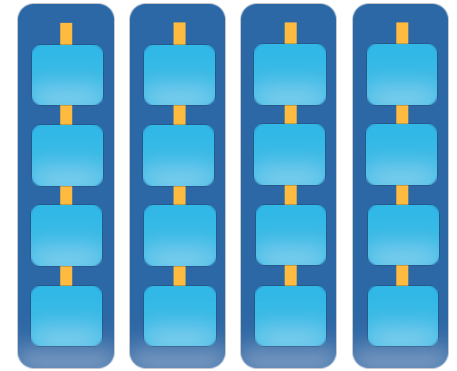
# Terms of Confusion:

## “job”



*Can be used to describe:*

- a program and what it does
- an item in the queue
- all items of the same HTCondor “cluster”
- an entire batch of submit files
- an entire multi-step workflow (e.g. a DAG)



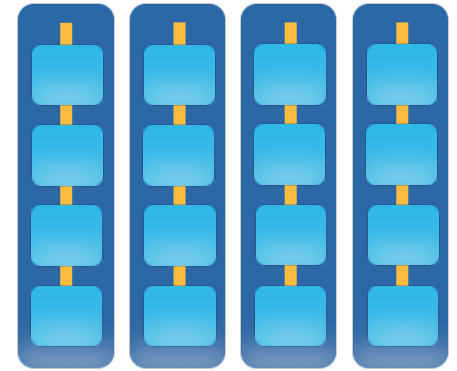
# Terms of Confusion:

## “job”



*Can be used to describe:*

- a program and what it does
- an item in the queue**
- all items of the same HTCondor “cluster”
- an entire batch of submit files
- an entire multi-step workflow (e.g. a DAG)



# Terms of Confusion:

## **“job” – define terms**

“Referring to each submit file as a ‘batch’, and each queued process as a ‘job’ ...”

“Since the first node of your DAG submits 10 jobs ...”

# Terms of Confusion:

**“cluster” – define terms**

HTCondor \$(Cluster)

-OR-

an organized set of hardware?



# Terms of Confusion:

**acronyms, abbrev'ns, and jargon**

**FTP**

**wget**

**TCP**

**LDAP**

**startd**

**DNS**

**schedd**

**SL6**

# Terms of Confusion:

**acronyms, abbrev'ns, and jargon**  
**- use general terms**

“... download the file using a ‘wget’  
command.”

“The version of Linux we use (Scientific  
Linux 6, or “SL6”) ...”

# Terms of Confusion:

other



log  
node  
process  
workflow

# Answer the REAL Question



**Identify questions that indicate confusion.**

Is the user asking the *wrong* question?

**Anticipate the next or ultimate question.**

Is the user on the way to bigger ideas?

**Focus on solutions and expectations.**

# In Summary ...



**Know your audience!** (Users are people)

**Implement the ‘right’ people as communicators.**

**Provide clear explanations, catered to the individual.**

**Be aware of “terms of confusion”.**

**Lead the user to their own expanded, accurate understanding.**

# And Finally ...



## Communicate about Communication!

Lauren Michael  
lmichael@wisc.edu

**chtc@cs.wisc.edu**