The End and the Beginning

Tim Cartwright

OSG Deputy Executive Director and OSG User School Director
University of Wisconsin–Madison
The End…
We must adjust to changing times and still hold to unchanging principles.

— Attributed to former U.S. President Jimmy Carter, who attributed it to his high school teacher, Julia Coleman
Why?
Gartner Hype Cycle

As of August 2021

Plateau will be reached: 〇 < 2 vrs. 〇 2–5 vrs. 〇 5–10 vrs. ▲ >10 vrs. ✗ Obsolete before plateau

Recent Paradigm Shifts

1970s  Computing capacity packaged and sold in small units

2000s  Computing capacity available to lease by the minute

**BUT:** Principles haven’t changed with these shifts!
# A Brief History of HTC

<table>
<thead>
<tr>
<th>Year</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>1983</td>
<td>Miron Livny completes Ph.D. thesis</td>
</tr>
<tr>
<td>1985</td>
<td>First Condor deployment</td>
</tr>
<tr>
<td>1992</td>
<td>Completed run of 250,000 jobs</td>
</tr>
<tr>
<td>1994</td>
<td>LHC approved</td>
</tr>
<tr>
<td>1996</td>
<td>Introduced “High Throughput Computing”</td>
</tr>
<tr>
<td>2000</td>
<td>Start of Trillium project (PPDG + GriPhyN + iVDGL)</td>
</tr>
<tr>
<td>2004</td>
<td>Start of EGEE (Enabling Grids for E-sciencE)</td>
</tr>
<tr>
<td>2005</td>
<td>Start of Open Science Grid</td>
</tr>
<tr>
<td>2009</td>
<td>LHC Run 1 begins</td>
</tr>
<tr>
<td>2010</td>
<td>Perspectives on Grid Computing</td>
</tr>
<tr>
<td>2010</td>
<td>Start of EGI (née European Grid Initiative)</td>
</tr>
<tr>
<td>2012</td>
<td>LHC detects Higgs boson</td>
</tr>
<tr>
<td>2015</td>
<td>LIGO detects first binary black hole merger</td>
</tr>
<tr>
<td>2017</td>
<td>First release of HTCondor Annex, to work with clouds</td>
</tr>
<tr>
<td>2022</td>
<td>PATh Facility deployed</td>
</tr>
</tbody>
</table>
Some Principles
#1: Unity and Autonomy
Two Principles of Distributed Computing

- **Unity**: Parts of system try to reach common goal
  - Goal determines rules that control independent parts
  - E.g., OSPool resources should always be running jobs

- **Autonomy**: Parts are autonomous but cooperate
  - One part can refuse a request from other part
  - E.g., a resource provider (site) can turn away pilots

- There are always trade-offs
Autonomy => Ownership

- In 1985, HTCondor added the idea of *resource ownership* as a key extension to prior work in distributed computing.
- Resources have owners, and those owners must have the ability to decide how their resources are used… or else!
HTC Is Sharing

• Should I share my resources and, if so, with whom and when?

• “HTC is about sharing across many jobs, many users, many servers, many sites, and (potentially) long-running workflows.”

– Miron Livny
“Now you have a community of customers who are motivated to share and act as consumers, providers, or both.”

– Miron Livny
#2: Automation
HTC Needs Automation

HTC requires **automation**, as it is a 24-7-365 activity that scales well beyond human interaction.

\[
\text{FLOPY} \neq \text{FLOPS} \times (60 \times 60 \times 24 \times 365)
\]

\[
1 \text{ job} \times 100 \text{ KHours} \neq 100K \text{ jobs} \times 1 \text{ Hr}
\]
Automation I: Get Resources & Run Jobs

**Central Manager**

- **negotiator**
- **collector**

**Submit Machine**

- **schedd**
- **shadow**

**Execute Machine**

- **startd**
- **starter**

1. submit job
2. request job details
3. send jobs
4. notify of match
5. claim
6. start
7. transfer exec, input
8. start
9. transfer output

send periodic updates
Automation II: User Workflows

Source: Professor Dane Morgan, University of Wisconsin–Madison
Automation III: System Maintenance

#3: Submit locally, run globally
OSG Is Getting There…

- Access Point
- Local Cluster
- OSG Metascheduling Service
  - Allocation
  - Sharing
  - Purchasing
- National Supercomputer
- Collaborator’s Cluster
- Nationally Shared Clusters
- Commercial Cloud
… But It Is Still Complex …
… and Takes a Lot of Care
#4: Focus on Users
HTCondor has been driven by what users (you!) need and expect, balanced with core principles

In contrast with “If you build it, they will come.”

Example: Edgar Spalding – his early work pushed our abilities to handle large input files (images)
Large Hadron Collider
IGWN (LIGO-Virgo-Kagra)
Event Horizon Telescope

Source: EHT Theory Working Group, CK Chan
Who Is Next?
#5: Teach To Fish
Give a person a fish and you feed them for a day.

Teach a person to fish and you feed them for a lifetime.
What Does “Teach to Fish” Mean for Us?

• Spurs innovation
  – Researchers who understand their own computing tools can see ways to innovate and make discoveries
  – For example, our Showcase speakers!

• Multiplies effort
  – By teaching researchers (you!) to solve computing challenges, we can help many people with few staff
  – And if you teach others (in your lab, etc.), our effort becomes exponentially powerful
Some Statistics on OSG Facilitation

- **375** new OSG Connect accounts in past year (each with initial consultation, at least)
- **80** tickets (support, ) in past quarter
- **26** institutions engaged in past quarter
- **20** visits to office hours in past quarter
- All with just **2.5 FTEs** over five people!
... And The Beginning
Getting Resources
Free Resources – In Your Lab

Server or cluster in your lab
👍 Not your laptop, control everything
👎 Buy and maintain it, not a lot of resources
Free Resources – Local Cluster

Department or campus cluster
👍 No/low direct costs, local help
👎 Shared; maybe Slurm, PBS/Torque, LSF…

No campus cluster? Talk to CIO!
Note! NSF CC* Compute awards
Free Resources – Collaborators

Collaborators
👍 No/low direct costs, may be tailored to project
👎 Shared, project-specific, hard to come by

https://www.dunescience.org/about-the-collaboration/
**Free Resources – Science Gateways**

**Science Gateways** (e.g., web front-end to a cluster)

👍 Easy to use, no/low cost

👎 Only for pre-defined use cases
• **Commercial clouds** *(Amazon, Google, Microsoft, …)*
  
  👍 Don’t own, high availability, many options (e.g., GPUs), …
  
  👎 Pay/hour, data out may be costly; challenging to manage

• **Managed clouds** *(Azure CycleCloud, Globus Genomics, …)*
  
  👍 As above, but *less* to manage
  
  👎 Costs more (paying someone to manage), fewer options?

• But keep commercial options in mind:
  
  – Credits may be available
  
  – May be able to write into grants
  
  – May be helpful for burst of activity (e.g., for a deadline)
The **PATh Facility** is a purpose-built, national-scale dHTC resource meant to deliver computational capacity to **NSF** researchers

- New in 2022!
- Can apply for credits – existing or new NSF award
- Credits go toward dedicated dHTC resources
- See [PATh website](https://www.nsf.gov/pubs/2022/nsf22051/nsf22051.jsp) and NSF DCL 22-051 for more:
Your School Accounts

• **learn.chtc.wisc.edu** – may keep for ~1 year
  – We will warn you before removing account

• **OSG Connect** account
  – If just created for School, temporary (~1 month)
  – But, just take easy steps to convert to full account
  – See School website:
    https://osg-htc.org/user-school-2022/logistics/projects/

• Remember that Access Points are not backed up!
Staying in Touch
How to Reach Us

• For OSG Connect, Connect Client, OSG sites…
  – user-support@opensciencegrid.org
  – Reaches the OSG Research Computing Facilitators

• For learn, CHTC, and anything else
  – user-school@opensciencegrid.org
  – Reaches Tim and Christina… and indirectly, many others

• Any time, for any reason, email us directly:
  – Tim Cartwright <cat@cs.wisc.edu>
  – Christina Koch <ckoch5@wisc.edu>
Online Support

Websites

- htcondor.org
- htcondor.org/manual
- www.opensciencegrid.org
- support.opensciencegrid.org
- osgconnect.net
- chtc.cs.wisc.edu
- chtc.cs.wisc.edu/guides

- HTCondor homepage
- HTCondor manual
- OSG homepage
- Forums, docs, support
- OSG Connect
- CHTC Website
- CHTC How-To Guides

Mailing Lists

- user-school@opensciencegrid.org
- help@opensciencegrid.org

will remain

general OSG help
Meetings

• OSG All-Hands Meeting
  – Traditionally in March, likely moving for 2023
  – David Swanson Award => former School participant

• HTCondor Week

• European HTCondor Workshop
Final Logistics
Travel

• Transportation emails went out today
  – Most people will take taxi/rideshare
  – Payer should write down who shared the ride

• Remember the travel advice page:
  https://osg-htc.org/user-school-2022/logistics/travel-advice/

• If you encounter issues:
  – Try the airline first
  – Call Travel Inc. if necessary (we are charged $18)
  – Let us know (if you want to and have time)
Reimbursements

• What can be reimbursed?
  – Bus fare to/from Madison (if you paid)
  – Taxi or rideshare to/from Madison airport
  – Driving mileage, *if prearranged*
  – Dinners (Sunday – Thursday), up to $30 each
  – Other *prearranged* things
  – It is best to have actual receipts

• How to get reimbursed?
  [https://osg-htc.org/user-school-2022/logistics/reimbursements/](https://osg-htc.org/user-school-2022/logistics/reimbursements/)
Short Essay (If No Lightning Talk)

• Similar to the lightning talks, but written
  – Basic summary of current research work
  – One key computational challenge
  – How will you apply new knowledge & skills to research?

• Due by the end of August

• Submit to School mailing list for review
Forward
Temper Hype With Principles

As of August 2021

Plateau will be reached:  ○ < 2 yrs.  ○ 2–5 yrs.  ● 5–10 yrs.  ▲ >10 yrs.  ❌ Obsolete before plateau
How to Work With Us

• We are driven by user needs and expectations, plus our principles

• So push us to help make your research possible

• And we may push on you to take your work even further!
Don’t let computing be a barrier to your science!
THANK YOU!
Closing Dinner

Union South, Industry Room and outside
(3rd floor)

6:30 p.m.
Acknowledgements

This work was supported by NSF grants MPS-1148698, OAC-1836650, and OAC-2030508