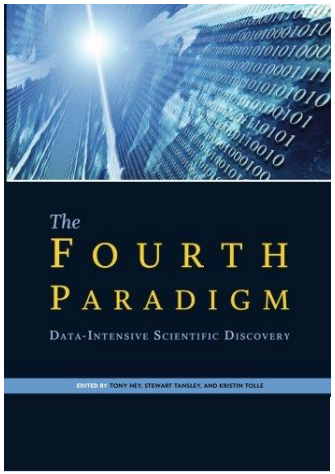
Using Clusters in Undergraduate Research

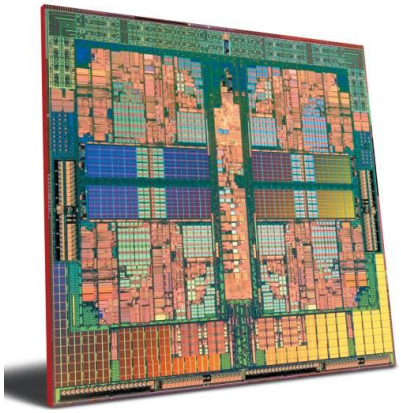
Peter Bui

University of Wisconsin - Eau Claire

Motivation

Problem: PDC Growth





NVIDIA CUDA

amazon web services™

XSEDE
Extreme Science and Engineering Discovery Environment

Parallel and **distributed computing** are becoming increasingly **important**.

Problem: HPC Education Adoption

Although we now have guidelines such as the **NSF/IEEE-TCPP PDC Curriculum** report, **integration** of such recommendations is still a **work in progress**…

Challenge: In place of classroom instruction (or in addition), how can we provide students an opportunity to **delve deep** into parallel and distributed computing?

Solution: Clusters + Research

Use clusters to introduce undergraduates to parallel and **distributed computing** in the context of:

- **Class Projects**
- **Independent Study**
- **Undergraduate Research**

} **Today's Focus**

Background

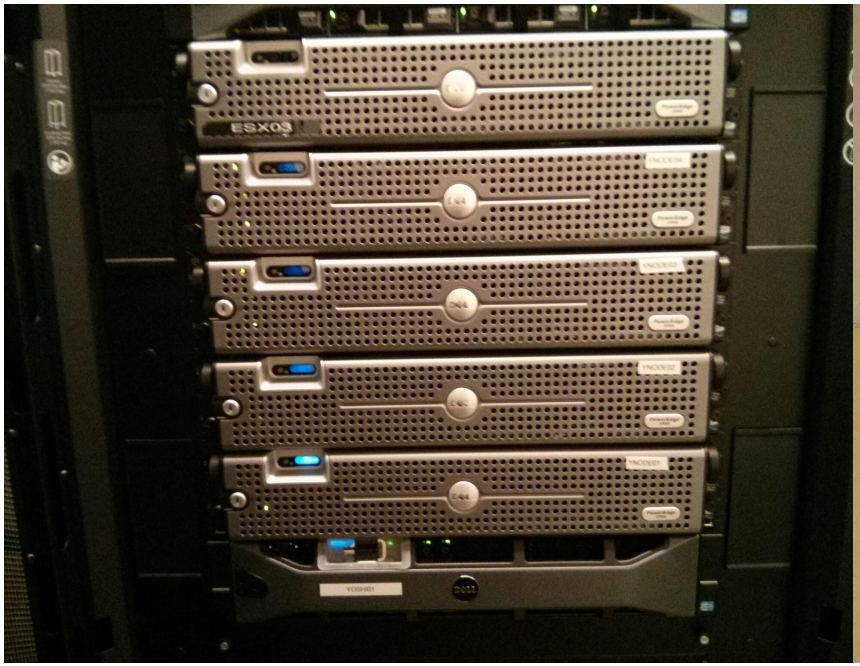
Undergraduate Research @ UWEC

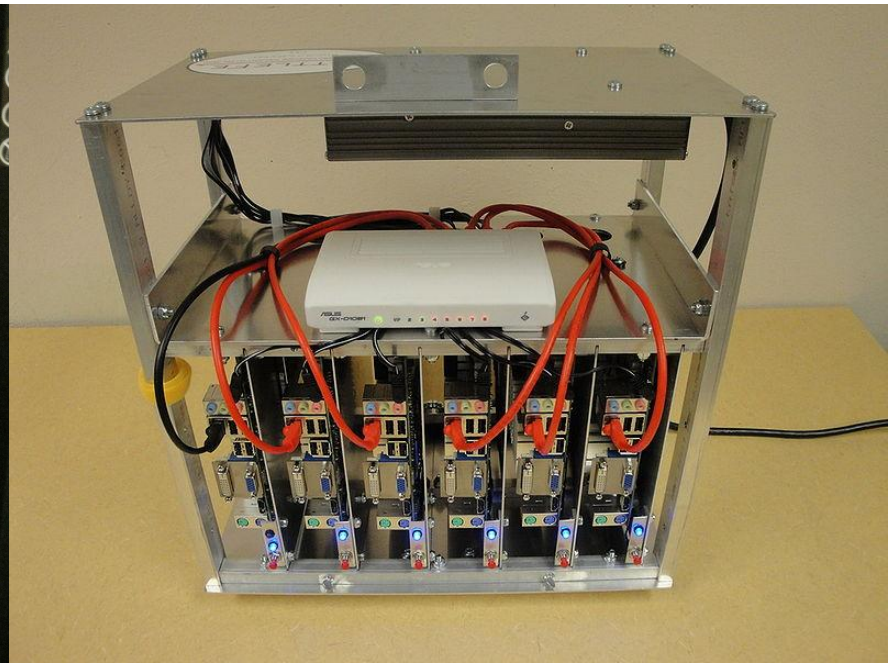
Students perform research:

- **Stipend**
- **Independent Study**
- **Hobby/Curious**

*As a mentor, my goal is to allow the students to **explore** their interests, **develop** skills, and **learn** more deeply about a particular topic.*

HPC @ UWEC

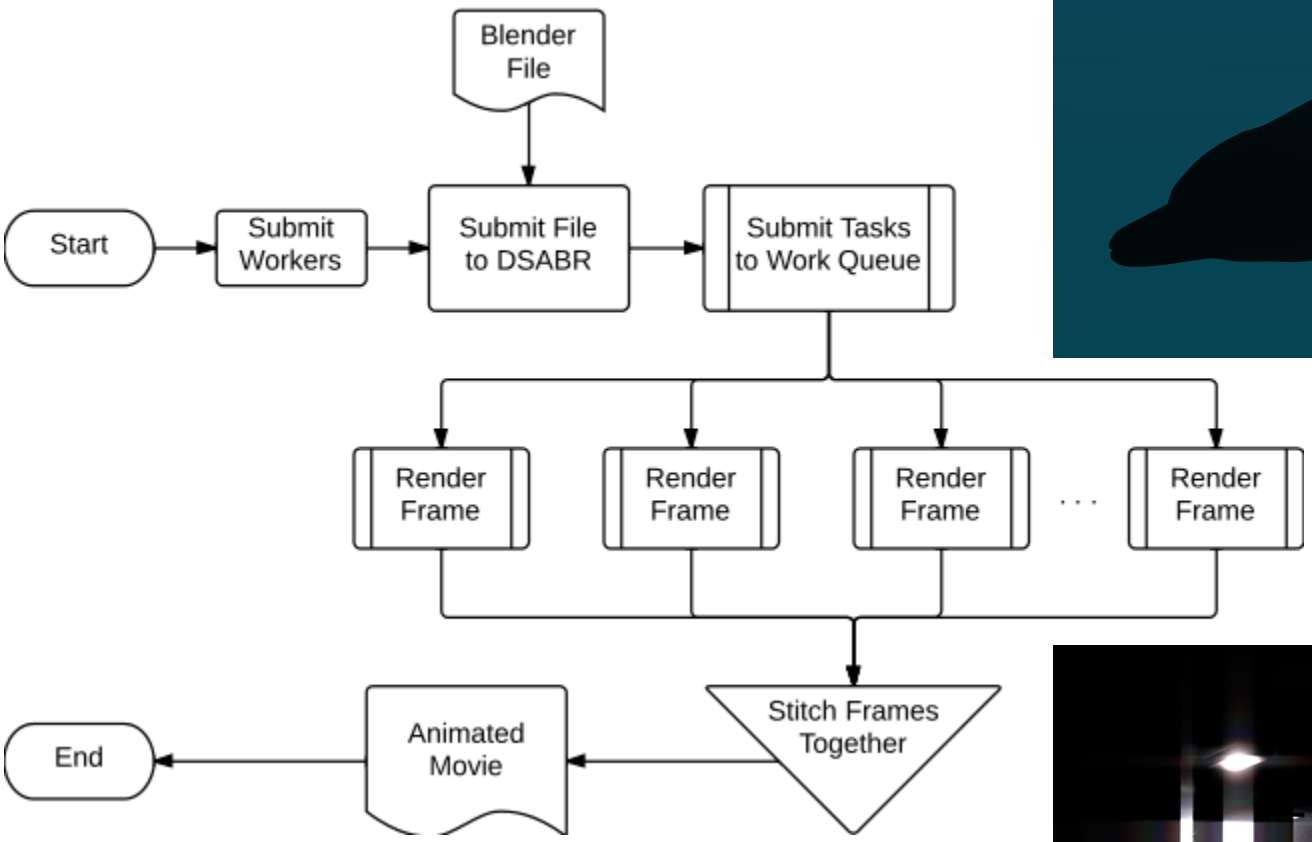


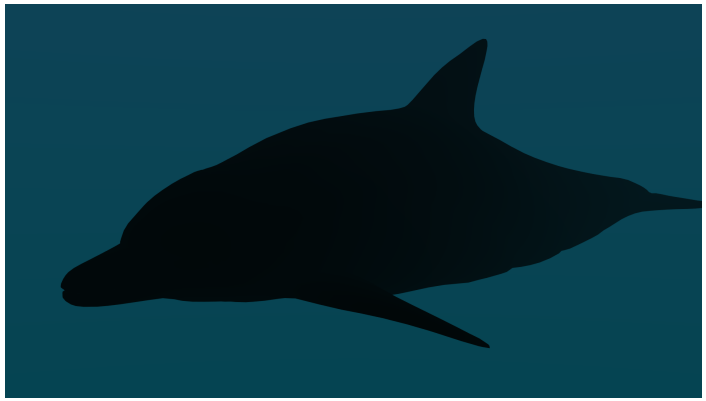


CENTER FOR
HIGH THROUGHPUT
COMPUTING

Undergraduate Research Projects

Distributed Animation Rendering





Distributed Animation Rendering

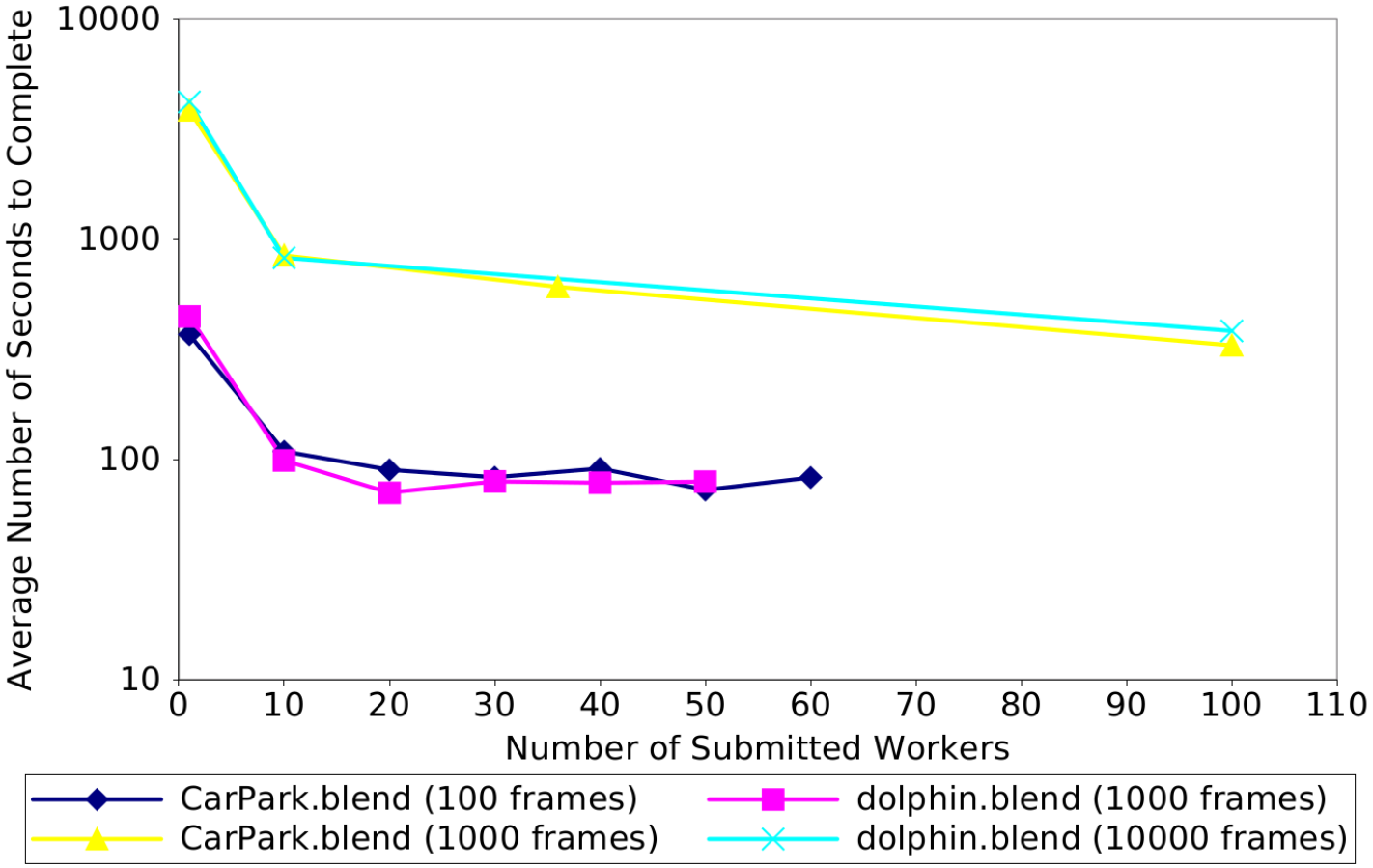


Photo Processing Pipeline

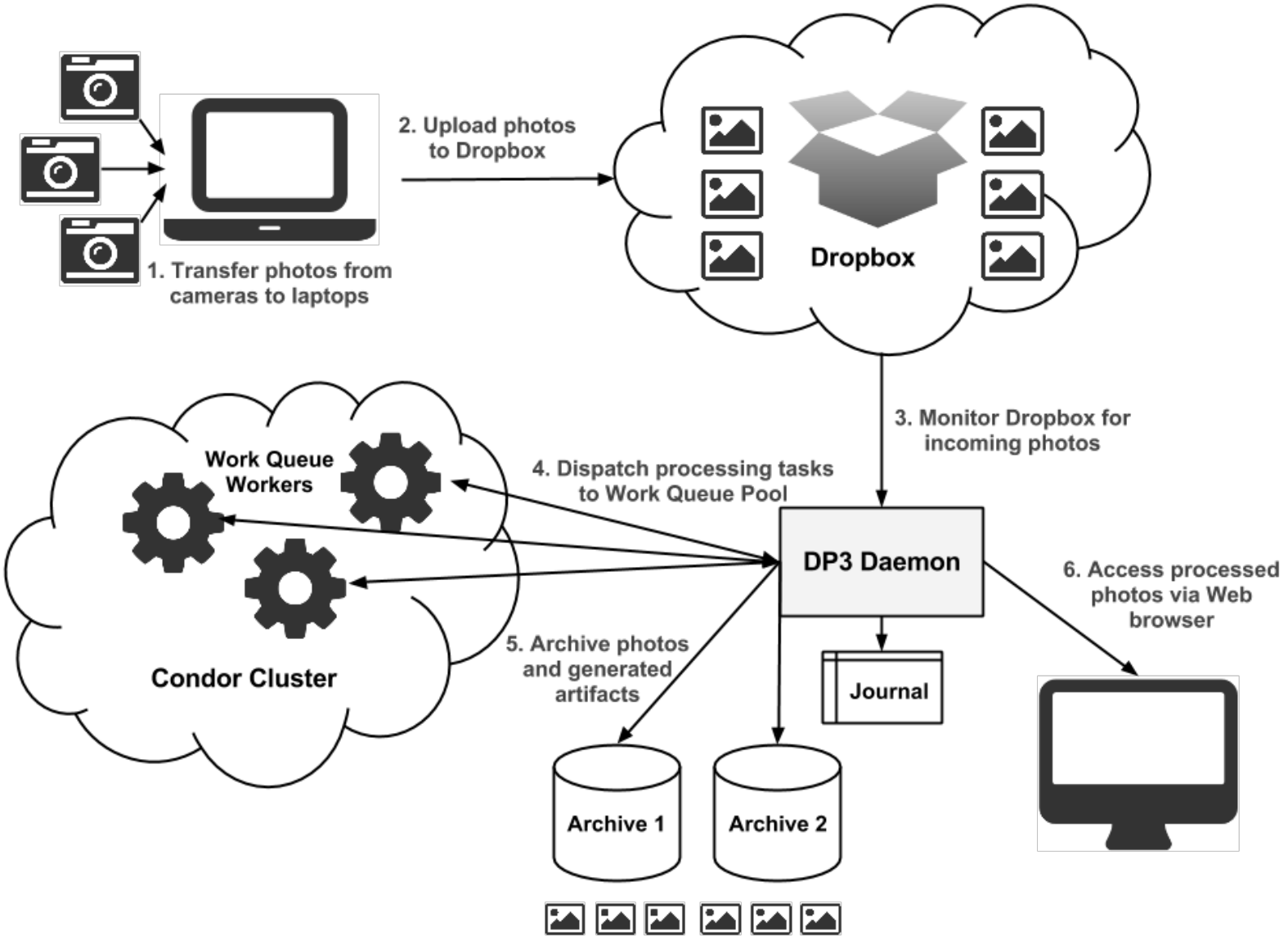
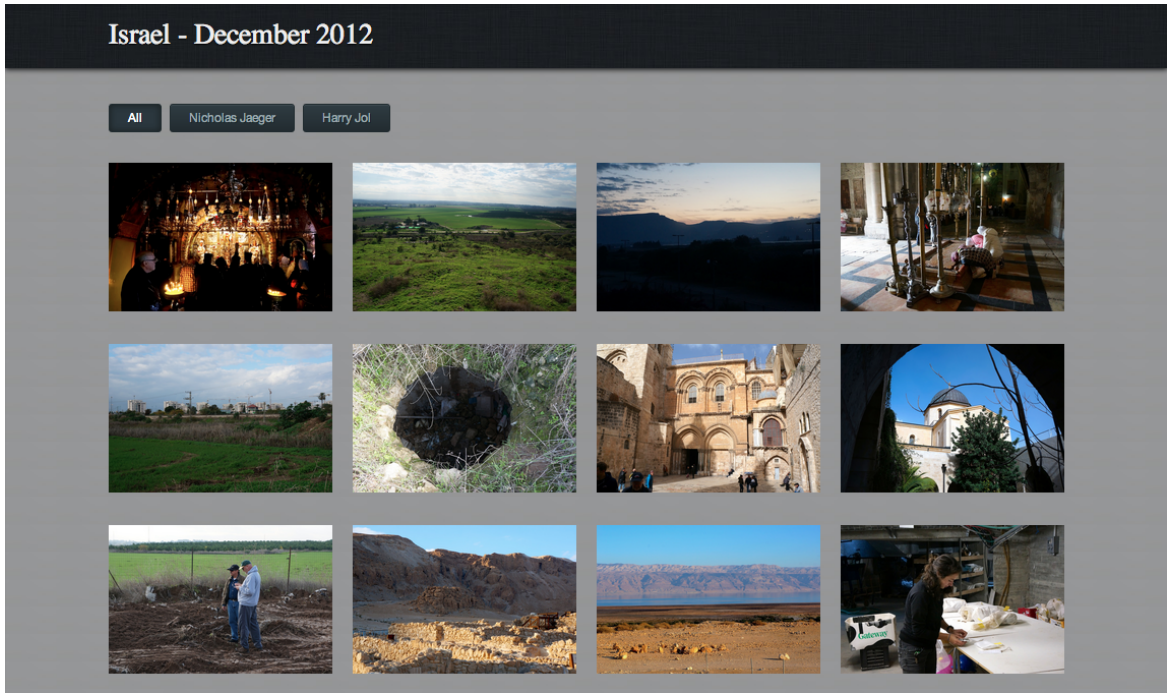
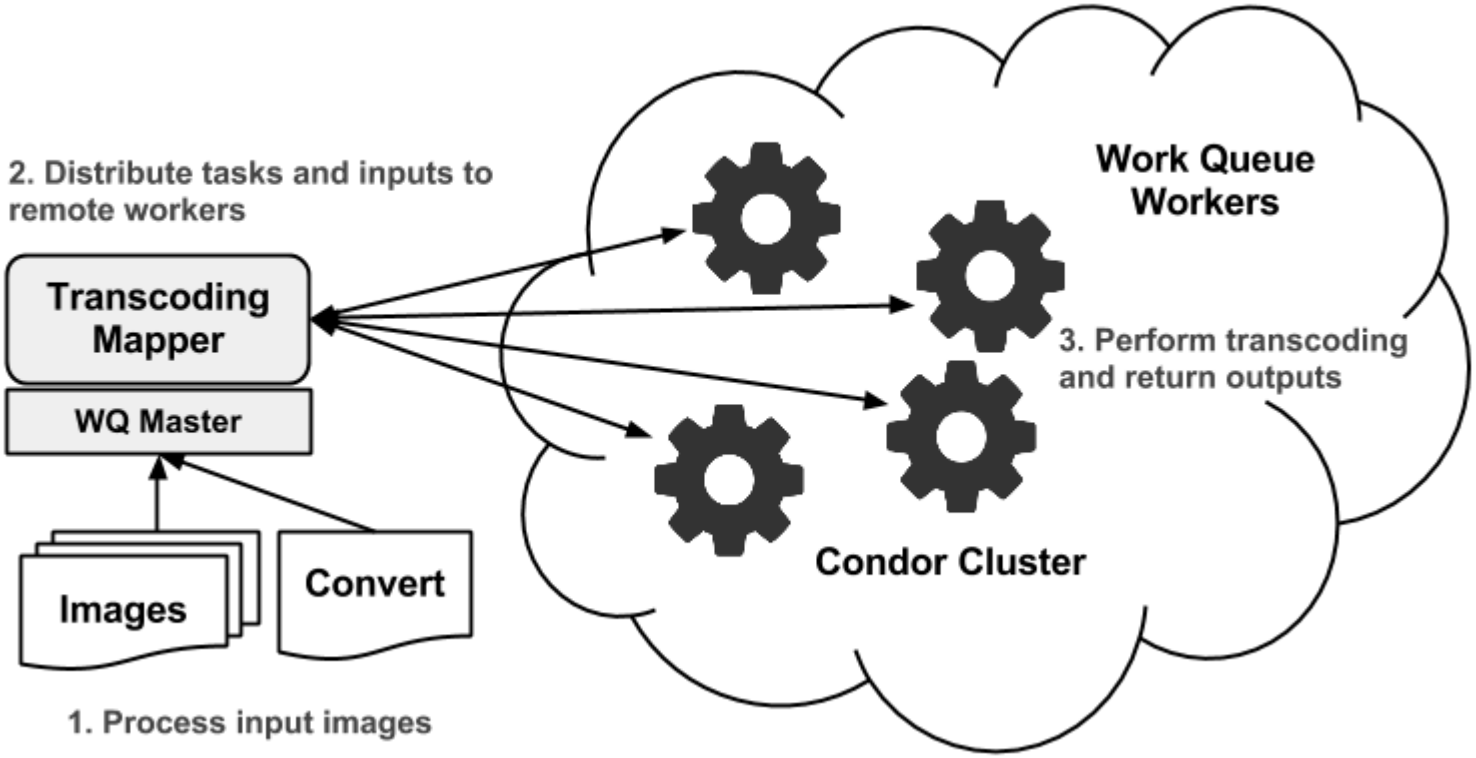


Photo Processing Pipeline



Num of Batches	448	**Num of Tasks Submitted**	7372
Max Batch Size	1385	**Num of Tasks Failed**	104
Min Batch Size	1		
Avg Batch Size	16.4		

Scalable Image Transcoding



Scalable Image Transcoding

File Size	Set Size	# of Workers						
		1	2	4	8	16	24	30
15KB	10	1x	1.47x	1.56x	2.13x	1.85x	2.00x	**2.40x**
	100	1x	1.60x	2.80x	4.43x	5.96x	6.42x	**6.44x**
	1000	1x	1.65x	3.12x	5.02x	7.97x	9.27x	**9.31x**
1MB	10	1x	1.65x	2.40x	2.78x	3.05x	3.73x	**3.87x**
	100	1x	2.10x	3.87x	6.55x	**9.56x**	7.65x	8.27x
	1000	1x	2.17x	4.28x	7.75x	11.2x	10.5x	**12.12x**
10MB	10	1x	1.84x	2.46x	2.88x	**4.48x**	3.43x	3.27x
	100	1x	1.98x	3.90x	4.95x	**7.34x**	4.61x	4.76x
	1000	1x	1.74x	3.97x	5.63x	**6.26x**	4.75x	4.93x

Lessons Learned

Applications, Not Infrastructure

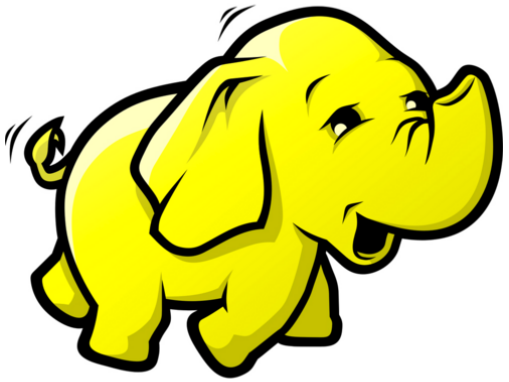
Because students lack distributed computing background:

*Focus on developing distributed **applications**, rather than low-level **infrastructure**.*

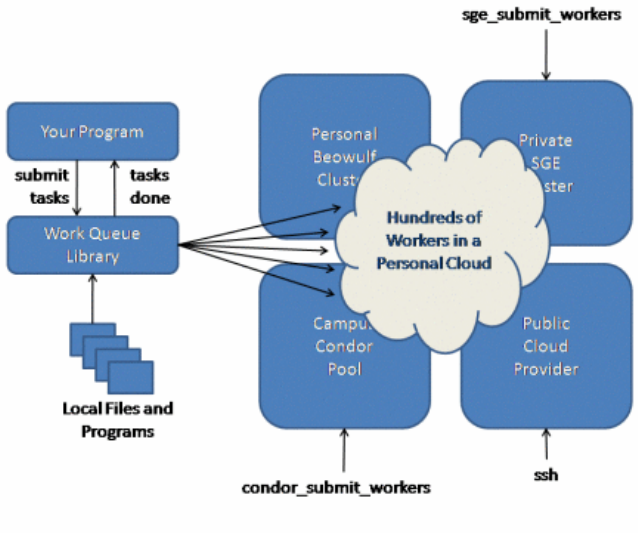
Goal: Allow students to **explore** distributed computing and to develop their interests and skills.

Utilize High-Level Frameworks

- Allow students to **dive quickly** into distributed computing
- Don't solve all of the **challenges** in distributed computing



*Frameworks are **scaffolding** that enables students to become **productive** in a **reasonable** amount of time*



Nothing is Straightforward

Programming distributed applications often yields **non-obvious** and **non-trivial** problems:

- **Identifying concurrency and parallelism**
- **Dependency management**
- **Local system limitations**
- **Network bottlenecks**

*Distributed computing is **hard**… but also **fun**!*

Practice Incremental Development

- **Iterative Development**

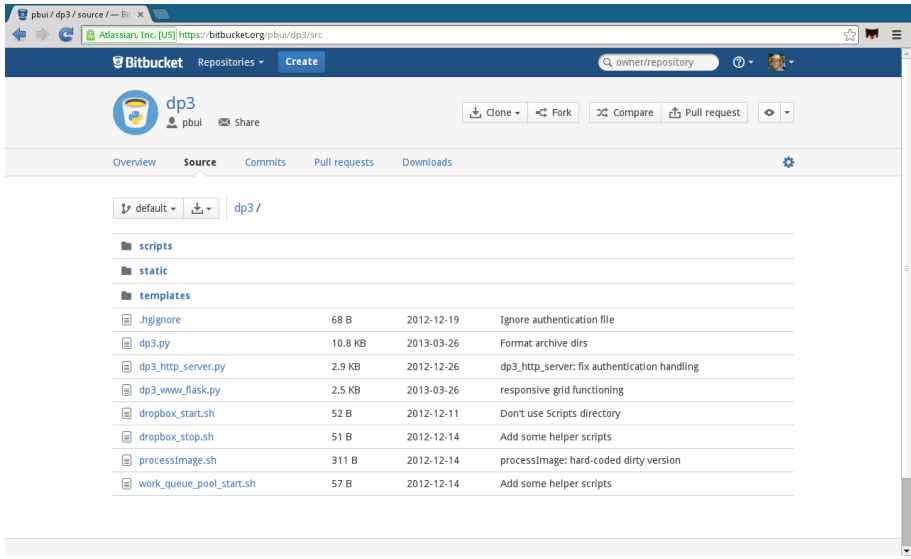
  Always have something that works

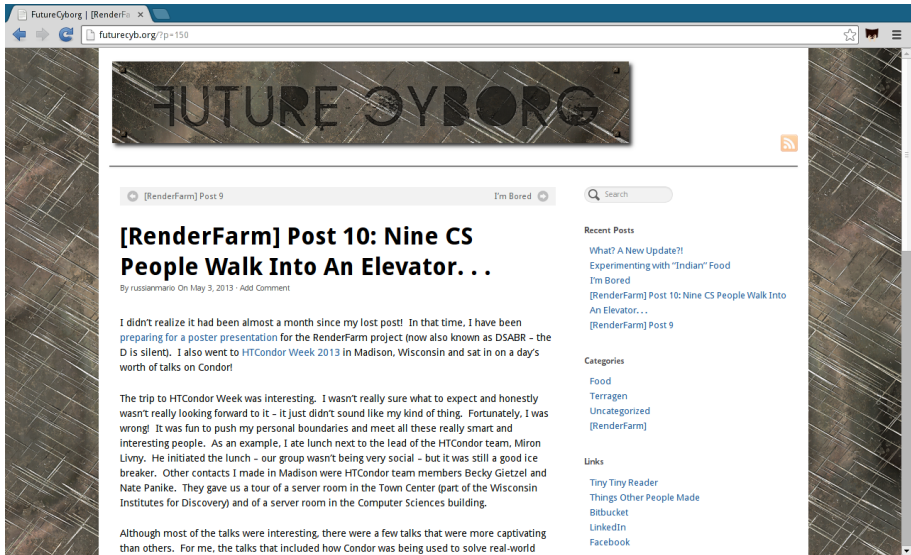
- **Version Control**

  Keep history of all work

- **Blogging**

  Describe and report progress

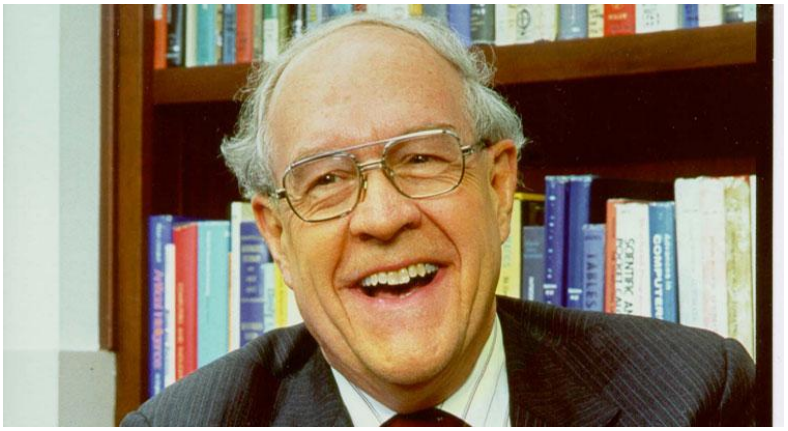




Seek Interdisciplinary Collaboration

"If the computer scientist is a toolsmith, and if our delight is to fashion power tools and amplifiers for minds, we must partner with those who will use our tools, those whose intelligences we hope to amplify."

- Fred Brooks



Internal and External Resources

- **UW - Eau Claire's ORSP**
  - Student stipends
  - Travel funding
  - Conference fees
  - Local presentation opportunities

- **Center for High Throughput Computing**
  - Condor flocking
  - Technical support
  - Student grants

Final Thoughts

Summary

Using clusters in **undergraduate research** is a great way to get students involved in **parallel and distributed computing**.

Distributed computing is **hard**…

- Focus on high-level **applications**
- Take advantage of **frameworks**
- Keep students **engaged** and **motivated**
- Reach out and take advantage of **resources**

Acknowledgements

- **Students**

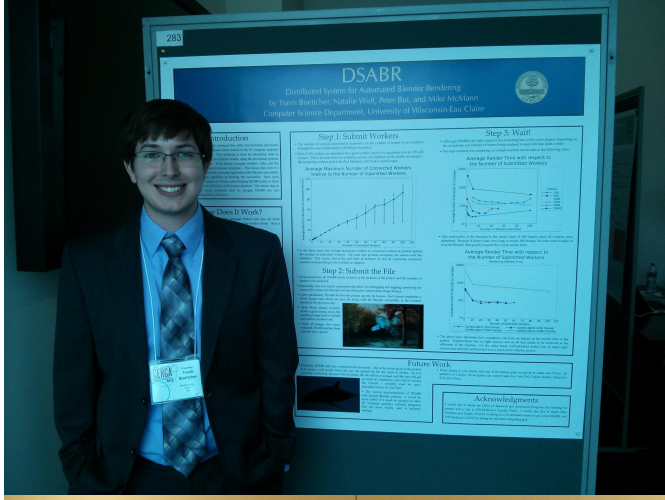
  Travis Boettcher, Nick Jaeger, Jeffrey Westphal

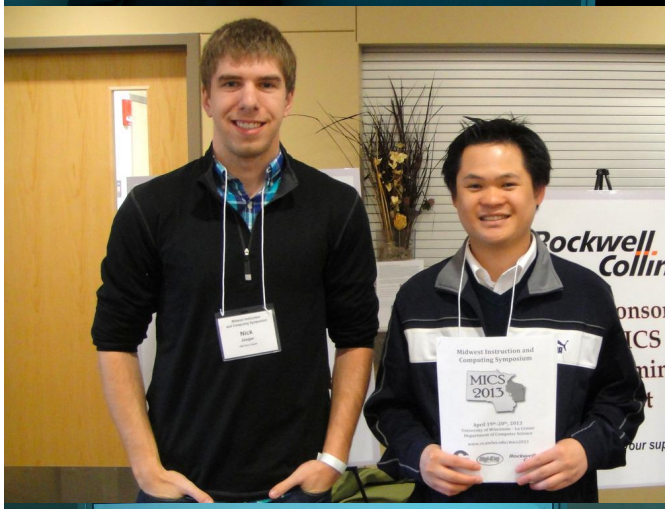
- **ORSP**

  Travel funding and student stipends

- **CHTC**

  HTCondor flocking





Questions?

Peter Bui

EMail: buipj@uwec.edu

WWW: http://cs.uwec.edu/~buipj

*The Computer Science department at UW - Eau Claire is **hiring two tenure-track positions** this Fall!*

Future Work

- **DSABR**
  - Support Maya
  - Take advantage of XSEDE and Amazon EC2
  - Web portal for art students

- **DP3**
  - Process photos, videos, and text
  - Generate blog posts
  - Use on more field trips (e.g. New Zealand)