How to do research?

Xiao Qin

Department of Computer Science and Software Engineering

Auburn University
http://www.eng.auburn.edu/~xqin
xqin@auburn.edu

Some slides are adapted from notes by Dr. Matthew Turk
My Story of Doing Research

• Undergraduate Student at Huazhong University of Science and Technology, (1996)
  – Real-Time Disk Scheduling in Unix
• M.S. Student at Huazhong University of Science and Technology, (1996-1999)
  – Real-Time and Fault-Tolerant Scheduling
• A doctoral student at the University of Nebraska-Lincoln (2001-2004)
  – I/O-Aware Load Balancing
• Assistant Professor at New Mexico Tech (2004-2007)
  – Security-Aware Scheduling
• Assistant Professor at Auburn University (2007-2010)
  – Energy-Efficient Storage Systems
• Associate Professor at Auburn University (2010-now)
  – Active Storage Systems
Caveat emptor

- These are *my opinions*, not departmental policies
- Talk to others to get *their* views

- These comments are intended for those who want to do research
  - All undergraduate students, MS students, and PhD students doing theses and projects
Why are you here?
Possible Reasons

**No Research Topic**
I couldn’t find a research topic

**What to do?**
I don’t know what I want to do as a graduate student

**Make Money**
I want to make money.

**Find Jobs**
I want to secure a good job.
Why I hope you’re here
Better Reasons

**Passion**
I am passionate about research, about engineering and their applications.

**Exploring**
I want to explore new intellectual territory and push the frontiers of technology

**Changing**
I want to change the world.

**Expert**
I want to become a world expert in XYZ.
Your Future Career


Phone and On-site Interviews

Programming skills
- Quickly learn a new programming language
- Programming experience

Problem solving skills

Personality
What Is A Typical Career Path?

**College**
- Earn college degree in CSSE

**Practice**
- Working for a company writing code
- Advance to analyst designer

**Management**
- An executive such as CIO
- A manager for large projects
- A system architect on increasingly large projects
- Start your own company
Motivation and Experience

Motivation: 0, 1, 2
Experience: 0, 1, 2

3 Lead
2 Management
4 Dedicate
1 Micro Management
A Success Story

• Undergraduate Research Assistant, 2005
• Adam Manzanares, Ph.D. May 2010
How to be a successful research assistant?

10 pieces of advice guaranteed to make you a successful research assistant.
1. Manage Yourself

- **Goals, priorities, and planning**
  - Set **goals**, and keep them updated
  - Make a **plan** for each day, week, month, quarter
    - “Failing to plan means planning to fail”
  - **Prioritize** – do important things first
  - Don’t waste time – kill your TV, xBox
    - Keep track of how you spend your time
    - Computer Science ≠ Web Browsing Engineering
    - “Is this activity helping me to achieve my REU goal?”
  - Keep a **notebook**, write these things down
Example 1 –
Keep Track of Your Time

<table>
<thead>
<tr>
<th>Work Total</th>
<th>Tasks</th>
<th>Time</th>
<th>Work Total</th>
<th>Tasks</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Week 11</td>
<td>7:55 AM</td>
<td>1:05</td>
<td></td>
<td>home</td>
<td>10:00</td>
</tr>
<tr>
<td></td>
<td>9:00 AM</td>
<td>0:20</td>
<td></td>
<td>HotStorage</td>
<td>10:30</td>
</tr>
<tr>
<td></td>
<td>9:20 AM</td>
<td>2:25</td>
<td></td>
<td>emails</td>
<td>12:30</td>
</tr>
<tr>
<td></td>
<td>11:45 AM</td>
<td>0:50</td>
<td></td>
<td>emails</td>
<td>12:30</td>
</tr>
<tr>
<td></td>
<td>comp2710</td>
<td>10:30</td>
<td></td>
<td>Lunch</td>
<td>12:57</td>
</tr>
<tr>
<td></td>
<td>12:35 PM</td>
<td>0:28</td>
<td></td>
<td>Lunch</td>
<td>12:57</td>
</tr>
<tr>
<td></td>
<td>1:03 PM</td>
<td>1:57</td>
<td></td>
<td>James</td>
<td>2:00</td>
</tr>
<tr>
<td></td>
<td>comp7500 Lab 2</td>
<td>2:30</td>
<td></td>
<td>Jash, Tom</td>
<td>3:30</td>
</tr>
<tr>
<td></td>
<td>Zhiyang</td>
<td>2:00</td>
<td></td>
<td>Jiong</td>
<td>5:25</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0:00</td>
<td></td>
<td></td>
<td>0:00</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0:00</td>
<td></td>
<td></td>
<td>0:00</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>9:35</td>
<td>Total</td>
<td>9:30</td>
<td></td>
</tr>
<tr>
<td>Work Total</td>
<td></td>
<td>9:07</td>
<td>Work Total</td>
<td>9:03</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Example 2 –
Keep Track of Your Time: a better approach

myHours.com

Add time: Thursday, May 12, 2011

### Daily Work:

<table>
<thead>
<tr>
<th>Project</th>
<th>Task</th>
<th>Start</th>
<th>Finish</th>
<th>Duration</th>
<th>Add. costs</th>
<th>Note</th>
<th>Edit</th>
<th>Delete</th>
</tr>
</thead>
<tbody>
<tr>
<td>Auburn</td>
<td>Research - Proposal</td>
<td>08:00</td>
<td>09:34</td>
<td>1:34</td>
<td>$0.00</td>
<td></td>
<td>Edit</td>
<td>Delete</td>
</tr>
<tr>
<td>Auburn</td>
<td>Service - Review Papers</td>
<td>09:34</td>
<td>10:41</td>
<td>1:07</td>
<td>$0.00</td>
<td></td>
<td>Edit</td>
<td>Delete</td>
</tr>
<tr>
<td>Auburn</td>
<td>Research - Career Development</td>
<td>10:41</td>
<td>11:00</td>
<td>0:19</td>
<td>$0.00</td>
<td></td>
<td>Edit</td>
<td>Delete</td>
</tr>
<tr>
<td>Auburn</td>
<td>Research - Proposal</td>
<td>11:00</td>
<td>12:44</td>
<td>1:44</td>
<td>$0.00</td>
<td></td>
<td>Edit</td>
<td>Delete</td>
</tr>
<tr>
<td>Auburn</td>
<td>Others - Lunch</td>
<td>12:44</td>
<td>13:17</td>
<td>0:33</td>
<td>$0.00</td>
<td></td>
<td>Edit</td>
<td>Delete</td>
</tr>
<tr>
<td>Auburn</td>
<td>Students - Men</td>
<td>13:17</td>
<td>14:50</td>
<td>1:33</td>
<td>$0.00</td>
<td></td>
<td>Edit</td>
<td>Delete</td>
</tr>
<tr>
<td>Auburn</td>
<td>Students - Ji</td>
<td>14:50</td>
<td>15:16</td>
<td>0:26</td>
<td>$0.00</td>
<td></td>
<td>Edit</td>
<td>Delete</td>
</tr>
<tr>
<td>Auburn</td>
<td>Students - Yixian</td>
<td>15:16</td>
<td>16:06</td>
<td>0:52</td>
<td>$0.00</td>
<td></td>
<td>Edit</td>
<td>Delete</td>
</tr>
<tr>
<td>Auburn</td>
<td>Research - Career Development</td>
<td>16:06</td>
<td>16:53</td>
<td>0:45</td>
<td>$0.00</td>
<td></td>
<td>Edit</td>
<td>Delete</td>
</tr>
<tr>
<td>Auburn</td>
<td>Research - Proposal</td>
<td>16:53</td>
<td>18:15</td>
<td>1:22</td>
<td>$0.00</td>
<td></td>
<td>Edit</td>
<td>Delete</td>
</tr>
</tbody>
</table>

Hours total: 10:15
Example 3: How to reply emails?

- Google: “How to Read 100 Emails, Fast”
- Check email once a day
- Group emails
- Reply to all the short emails - first with "yes" or "no" as an answer
- Write brief emails
- Long emails -> tasks -> must be prioritized
2. Develop Intellectual Discipline

- **Think!**
  - Set aside time for thinking. Really.

- **Read! (To be covered in another training session)**
  - Get to know the literature in your area intimately (not superficially)

- **Act!**
  - Don’t feel like you have to know everything first
  - Don’t worry about being wrong

- **Evaluate!**
  - Solicit feedback – most ideas aren’t so good…
Example 4 – Keep a notebook


* This paper analyzes eager techniques using three key parameters discuss their requirements, capabilities, and cost.


* Our research in this field can be done in three steps:
  1. Develop a theoretical framework including a series of replica control protocols
  2. Evaluate our replica control

3. Be proactive

• Don’t wait to be told what to do
  – Don’t be passive; in fact, be aggressive!
  – Make things happen

• You will not be spoon-fed
  – What you get out of the research program is a non-linear function of what effort you put into it.

• Research activities can be very unstructured
  – Unlike undergraduate studies
  – So it’s up to you (not your advisor)
11/18/2010
Completed Items:
1. Preliminary result shows we can benefit from prefetching, especially for the small files.
   if had not work, we can benefit from prefetching by 20% improvement.
   if had does work, we also can benefit from prefetching by 10%.

Next steps:
1. Experiment design, considering the elements:
   - single machine/cluster
   - changing the memory size
   - change the data size
   - single machine: 500MB 1G 2G 3G 4G
   - cluster: 10G 20G 30G 40G
   - application: Grep, wordcount, sort
   - large/small file

2. and then collect the CPU and IO utilization to explain why hadoop can benefit from prefetching. Is that possible to record the time when the data is available for execution.

10/13/2010
Completed Items:
1. Comparing the execution time of many small files with big file.

Next steps:
1. Consider the influence of prefetching rate and how to measure it.
2. Journal paper for the HCV'09 paper:
   - Conduct experiments using the new HP cluster.
   - Study network communications.
Example 6 – Dropbox to share document
4. Learn to communicate well

- **Speaking**
  - Communicate clearly

- **Writing**
  - Organization and clarity

- **Presenting**
  - Not just “talking,” but *communicating*
  - Even a lecture is a two-way interaction

- These are skills that can be learned!
  - Practice talks (videotaped), write short papers, ask friends and colleagues to help you, …
5. Develop an intellectual community

- Among your peers at Auburn, create something different and special
  - Ask questions
  - Discuss ideas
  - Brainstorm
  - Argue, challenge
  - Collaborate
6. Networking

- Get to know the people in the department (faculty and grad students), and other people in your field
  - Don’t wait – introduce yourself!
- Go to conferences and meet other REU students and “famous” researchers
  - Be aggressive!
- Talk with visitors: “pick their pockets”
  - You never know who will someday offer you a job, write a reference letter, review your paper, give you invaluable feedback or insight….
7. Choose a good research problem

- This is the hardest, and most important, part of research!
- The Goldilocks problem:
  - Not too hard, not too soft, not too hot, not too cold, not too big, not too small
- Think, read, act, evaluate
  - And talk to everyone – not only your advisor
- Passion or duty?
An Example
Data Placement in Hadoop Clusters
An Example
Data Placement in Hadoop Clusters (cont.)

- The MapReduce programming model is growing in popularity
- Hadoop is used by Yahoo, Facebook, Amazon.
Another Example – How to think
I/O Performance Bottleneck Problems in
Bioinformatics Applications
8. Understand the faculty

- We are very busy.

---

- We are very busy.
- That's no excuse. We do have time for you.
- We know more than you do.
- At least for a little while.
- But not as much more as you might think.
- We are not superior beings.
- Most of us have first names.
- Give us feedback too!
- We are part mentor, part colleague, part human.
9. Study successful people

- Senior grad students, faculty, pioneers, leaders in your field, ...
  - Read biographies
  - Who are your heroes, mentors?

- Seek advice
  - But modify it to your particular situation
10. Have a Life

- Work hard, networking, think, read, program, experiment, build, study, practice, ....
  - So little time and so much to do!!

- Still, amidst the chaos of the REU program, it is very important that you do not lose sight of who you are and what makes you tick.
  - Have a social life
  - Don’t neglect your family and friends, your health, your sanity
  - Do make time for things that are important and meaningful to you
Further Research

• Lots of links to good advice for graduate students:
  – http://www.cs.ucsb.edu/~mturk
  – Click on “Info for Students”
Further Research

http://www.eng.auburn.edu/~xqin

Xiao Qin, Associate Professor of Computer Science, Auburn University

Sep 3, 2010 ... UNL CSE graduate Xiao Qin received an NSF CAREER award in 2009 to investigate parallel disks that put substantial multicore computing power ...

www.eng.auburn.edu/~xqin/ - Cached - Similar
My webpage

http://www.eng.auburn.edu/~xqin
Xiao Qin's Presentations

Presentations

- Current Projects
- Completed Projects
- Research Funding
- Research Sponsors
- Presentations

Download our slides from SLIDESHARE PRESENTATIONS

1. Energy-Efficient Data Storage Systems. [PPT | Slideshare]
2. HDFS-HC: A Data Placement Module for Heterogeneous Hadoop Clusters. Note: This presentation is based on our paper - Improving MapReduce Performance via Data Placement in Heterogeneous Hadoop Clusters published in Proc. 19th Intl Heterogeneity in Computing Workshop, Atlanta, Georgia, April 2010. [PPT]
Summary

- How to do research?
- 10 pieces of advice
- Choose a good research problem

- Download the slides at
  
  http://www.slideshare.net/xqin74
Questions