How to do research?

AUBURN

UNIVERSITY

SAMUEL GINN
COLLEGE OF ENGINEERING

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

AUBURN UNIVERSITY
SAMUEL GINN COLLEGE OF ENGINEERING

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

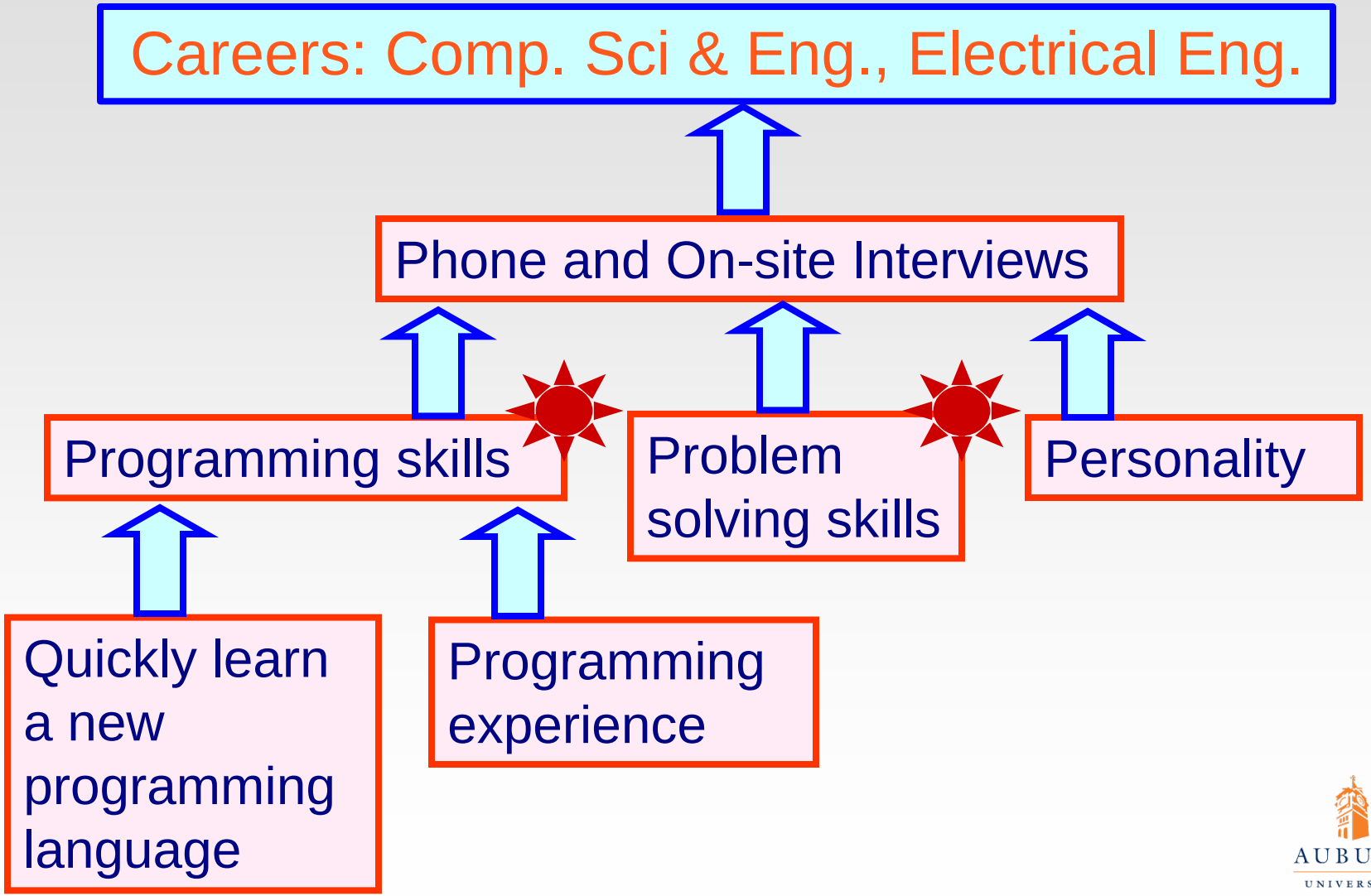
I want to change the world.

Expert

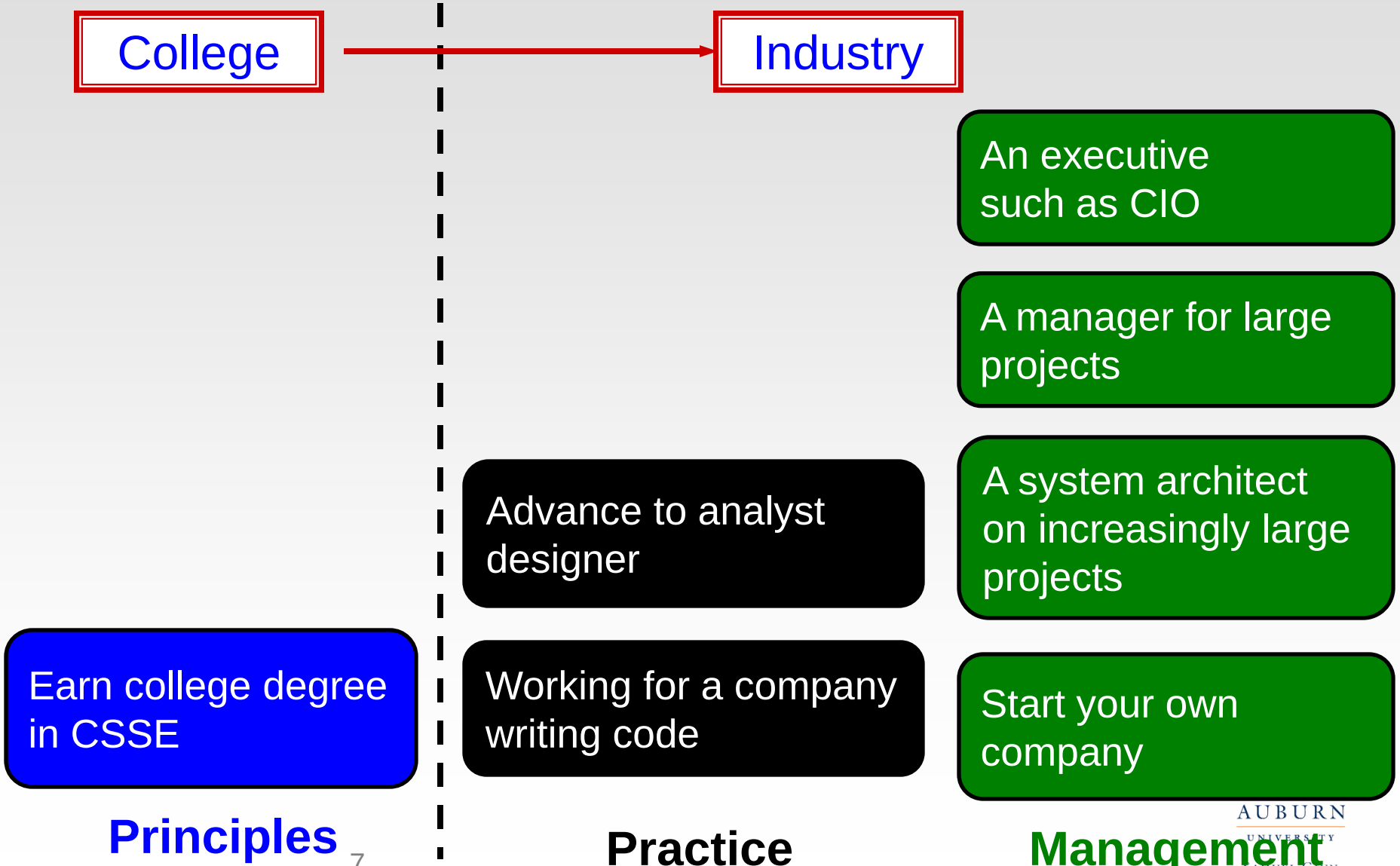
I want to become a world expert in *XYZ.*

AUBURN UNIVERSITY
SAMUEL GINN COLLEGE OF ENGINEERING

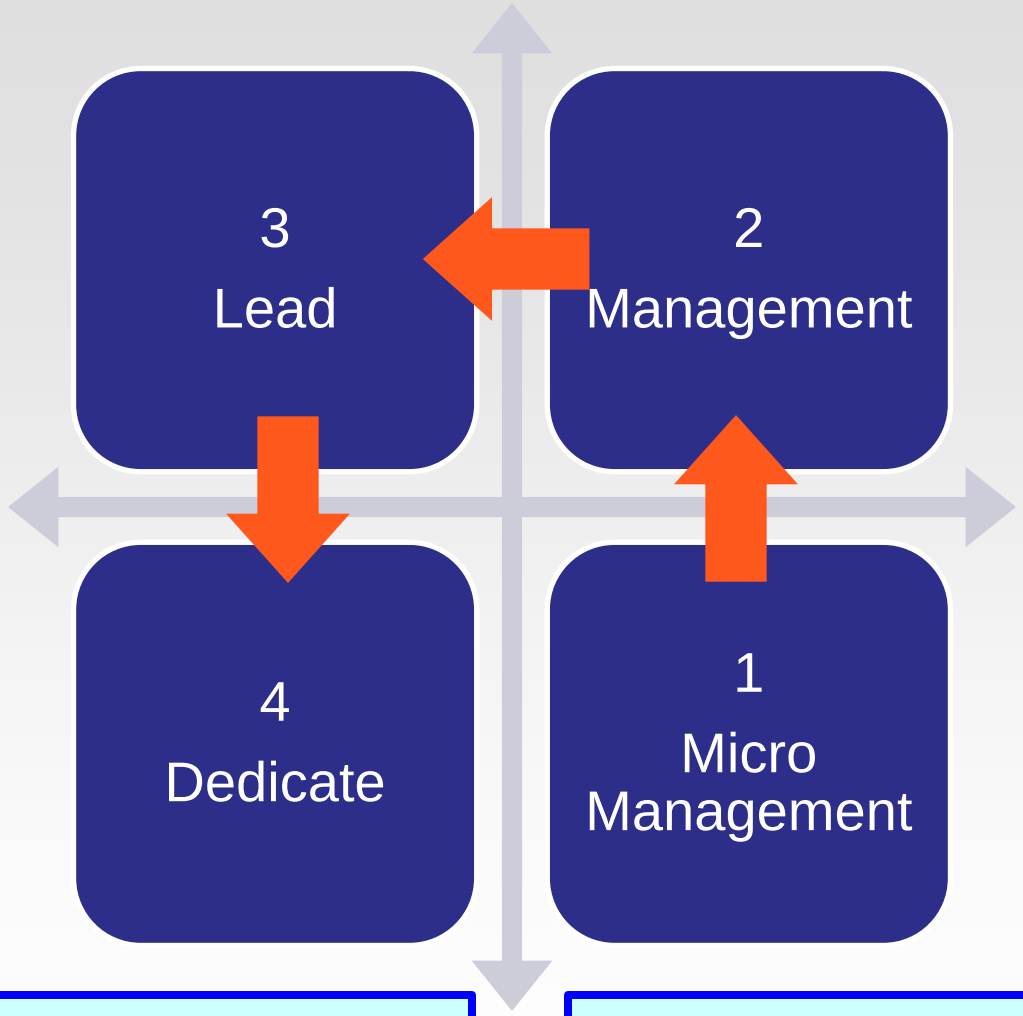
Your Future Career



What Is A Typical Career Path?



Motivation and Experience



Motivation: 0, 1, 2

Experience: 0, 1, 2

A Success Story

- Undergraduate Research Assistant, 2005
- Adam Manzanares, Ph.D. May 2010

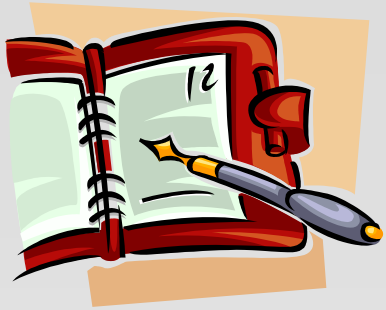
AUBURN UNIVERSITY

SAMUEL GINN COLLEGE OF ENGINEERING

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

Work Total			8:43	Work Total			8:40	Wo
3/22/2010	Week 11	Tasks	Time	3/23/2010		Tasks	Time	3/
7:55 AM	9:00 AM	UAH-Rimbuse, Admiss	1:05	7:55 AM	9:01 AM	yun	1:06	
9:00 AM	9:20 AM	Email	0:20	9:01 AM	10:00 AM	home	0:59	1
9:20 AM	11:45 AM	comp2710	2:25	10:00 AM	10:30 AM	HotStorage	0:30	
11:45 AM	12:35 PM	comp7500 Lab 2	0:50	10:30 AM	12:30 PM	emails	2:00	
12:35 PM	1:03 PM	Lunch	0:28	12:30 PM	12:57 PM	Lunch	0:27	
1:03 PM	3:00 PM	comp7500	1:57	12:57 PM	2:00 PM	James	1:03	
3:00 PM	5:30 PM	Zhiyang	2:30	2:00 PM	3:30 PM	Jash, Tom	1:30	
			0:00	3:30 PM	5:25 PM	Jiong	1:55	
			0:00				0:00	
			0:00				0:00	
Total			9:35	Total			9:30	Tot
Work Total			9:07	Work Total			9:03	Wo
3/29/2010	Week 12	Tasks	Time	3/30/2010		Tasks	Time	3/

AUBURN UNIVERSITY

SAMUEL GINN COLLEGE OF ENGINEERING

Example 2 –
Keep Track of Your Time: a better approach

myHours.com

Log-out | MHlite | Support | Help

Daily View | Monthly View | Reports | Projects | Tasks | Preferences

Add time: **Thursday, May 12, 2011**

Date: Thursday, May 12, 2011
Project: -- select Project --
Task: -- select task --
Start: Now (+ -)
Finish: Now (+ -)
Duration: + -
Additional costs:
Note:
Submit

<<< < May 2011 > >>>

Sun	Mon	Tue	Wed	Thu	Fri	Sat
1	2	3	4	5	6	7
8	9	10	11	12	13	14
15	16	17	18	19	20	21
22	23	24	25	26	27	28
29	30	31				

Daily Work:

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

hours total: 10:15

AUBURN UNIVERSITY
SAMUEL GINN COLLEGE OF ENGINEERING

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

on the performance.

29. "Database Replication Techniques : ~~A~~ three Parameter Classification" SRDS 2000

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

30. "Database Replication for Clusters of workstation" Ph.D 2000. * Motivation for database replication

* 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

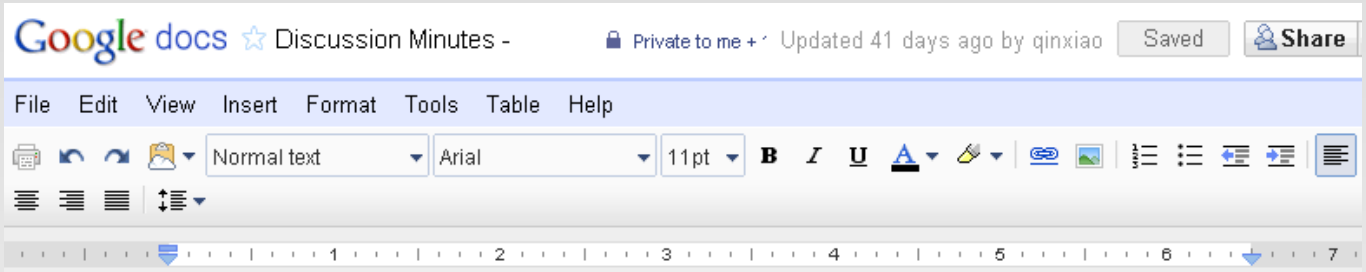
AUBURN UNIVERSITY

SAMUEL GINN COLLEGE OF ENGINEERING

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)

Example 5 – Discussion Minutes



11/18/2010

Completed Items:

1. Preliminary result shows we can benefit from prefetching. especially for the small files.
   if har does not work. we can benefit from prefecthing by 20% improvement.
   if har does work, we also can benefit from prefecthing by 10%.

Next steps:

1. Experiment design. considering the elements:
   - single machine/ cluster
   - chaging 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:

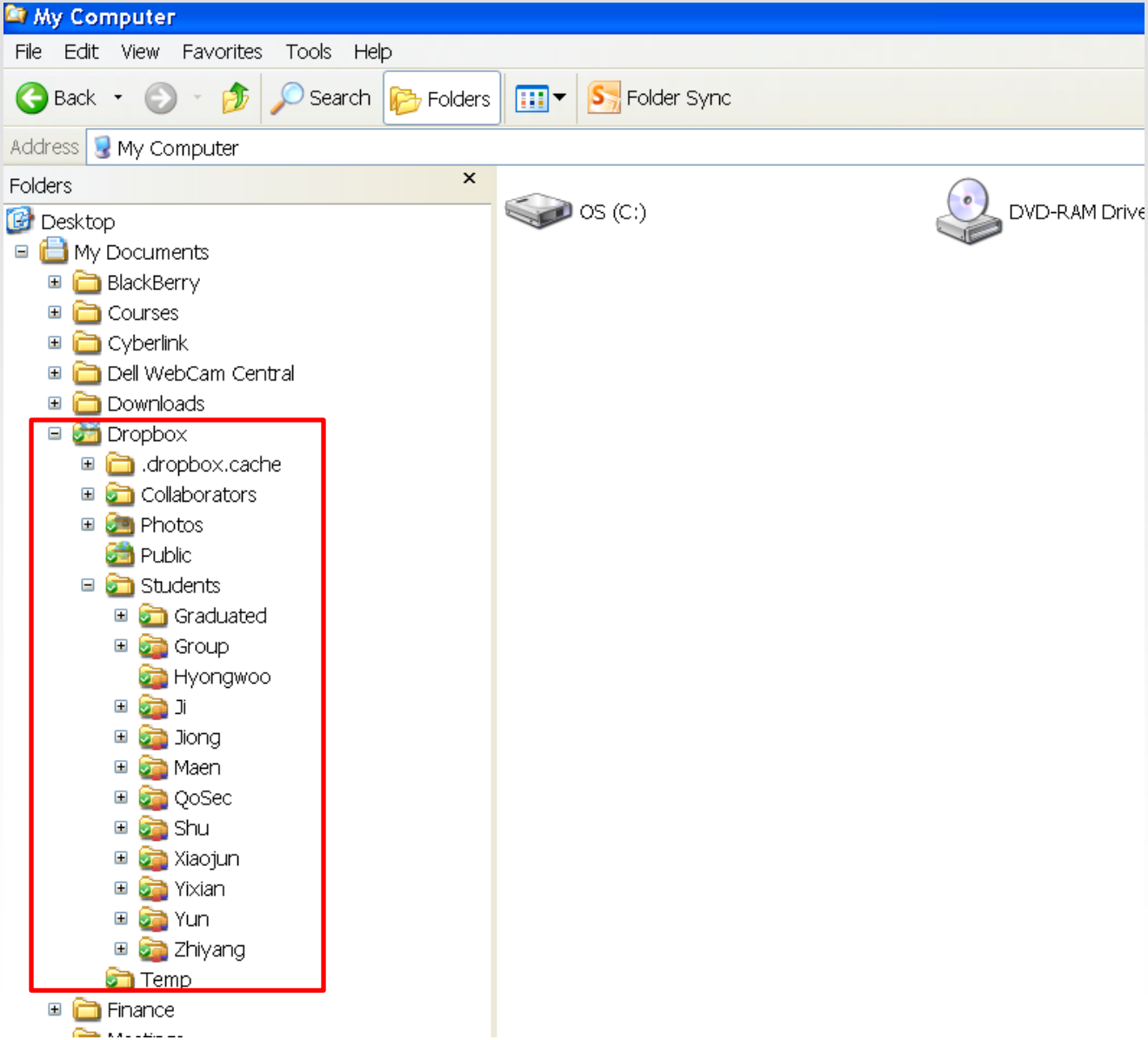
1. Comparing the execution time of many small files with big file.

Next steps:

1. Consider the influence of prefetching rate and how to measure it.
2. Journal paper for the HCW'09 paper.
   - Conduct experiments using the new HP cluster.
   - Study network communications.

AUBURN UNIVERSITY
SAMUEL GINN COLLEGE OF ENGINEERING

Example 6 – Dropbox to share document



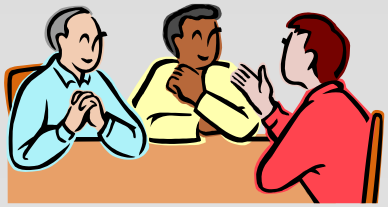
4. Learn to communicate well

- Speaking
  - Communicate clearly
- Writing
  - Organization and clarity
- Presenting
  - Not just “talking,” but *communicating*
  - Even a lecture is a two-way interaction

Your intelligence and ideas will be judged by your ability to communicate in English

- 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



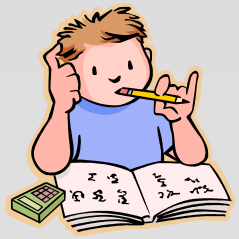
- Among your peers at Auburn, create something different and special
  - Ask questions
  - Discuss ideas
  - Brainstorm
  - Argue, challenge
  - Collaborate

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….

AUBURN UNIVERSITY
SAMUEL GINN COLLEGE OF ENGINEERING

7. Choose a good research problem



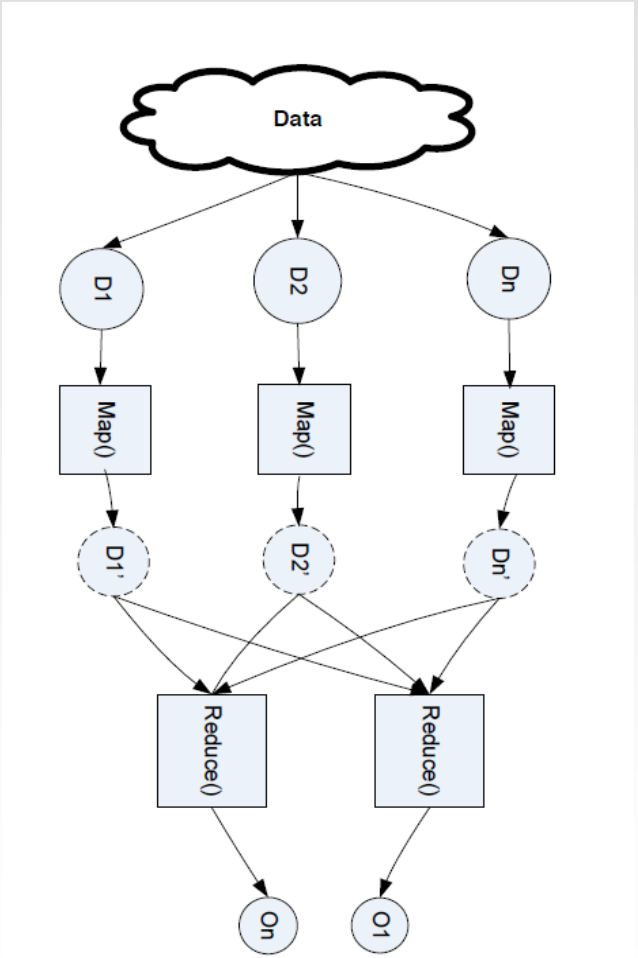
- This is the hardest, and most important, part of research!
- The Goldilocks problem:
  - Not too hard, not too soft, not too hot, not too cold, not too big, not too small
- Think, read, act, evaluate
  - And talk to everyone – not only your advisor
- Passion or duty?

An Example

Data Placement in Hadoop Clusters

An Example

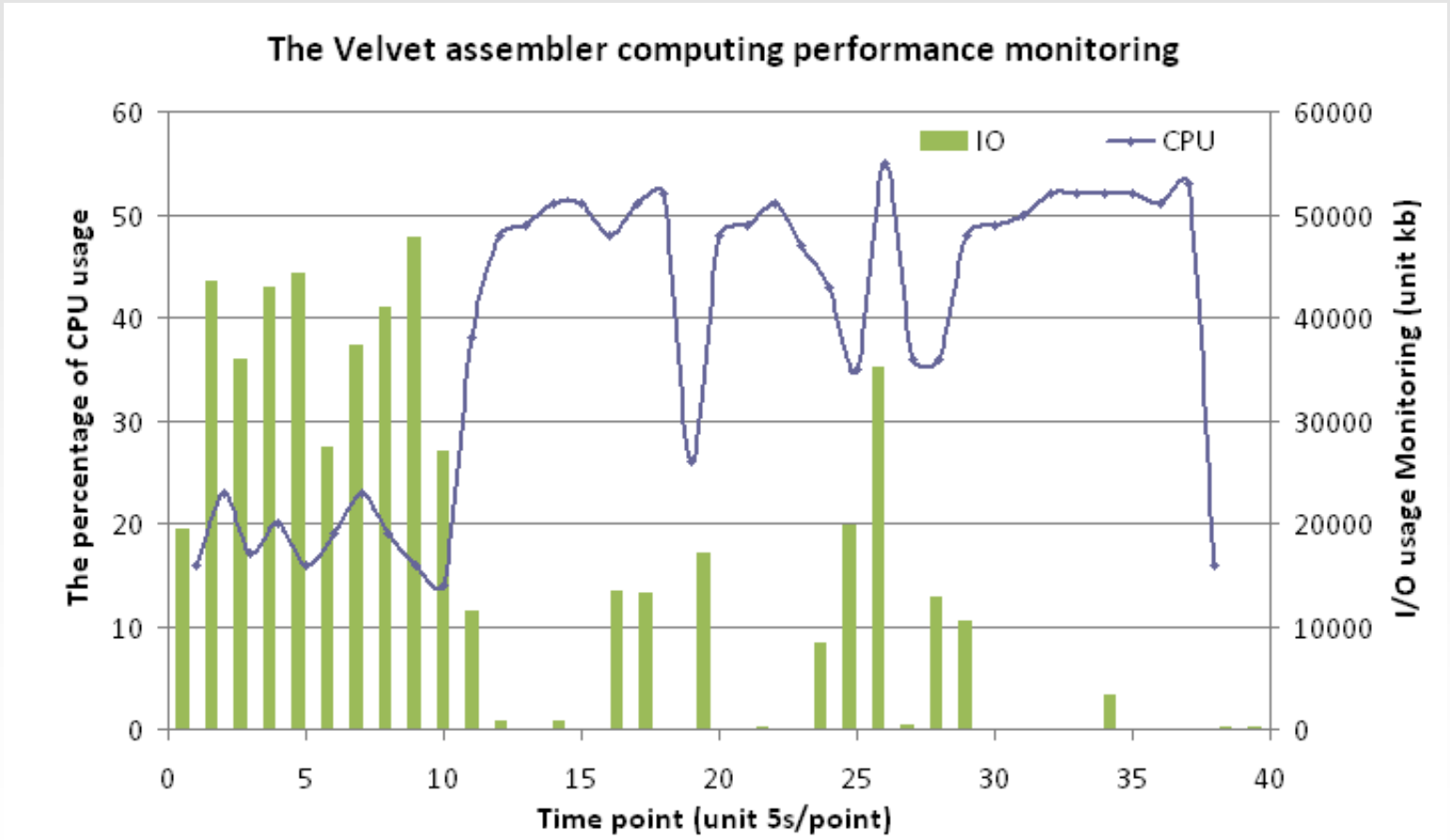
Data Placement in Hadoop Clusters (cont.)



- 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



AUBURN UNIVERSITY

SAMUEL GINN COLLEGE OF ENGINEERING

8. Understand the faculty

- We are very busy.

Apr 18 – 24 2010 Refresh Print Day

Sun 4/18	Mon 4/19	Tue 4/20	Wed 4/21	Thu 4/22	Fri 4/23
	8 – 9 Office Hour - comp2710		8 – 9 Office Hour - comp2710	8 – 5p Research Day	
	10 – 10:50 comp2710		10 – 10:50 comp2710	11:30 – Mais	10 – 10:50 comp2710
		1p – 2p James			
	2p – 2:50p comp7500	2p – 3p Tom and Josh	2p – 2:50p comp7500		2p – 2:50p comp7500
	3p – 4p Yun	3p – 4p Jiong	3p – 4p Shu	3:30p – 5p Group presentation	3p – 4p Zhiyang
	4p – 5p Office Hour - comp7500	4p – 5p Adam	4p – 5p Office Hour - comp7500		4p – 5p Xiaojun

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

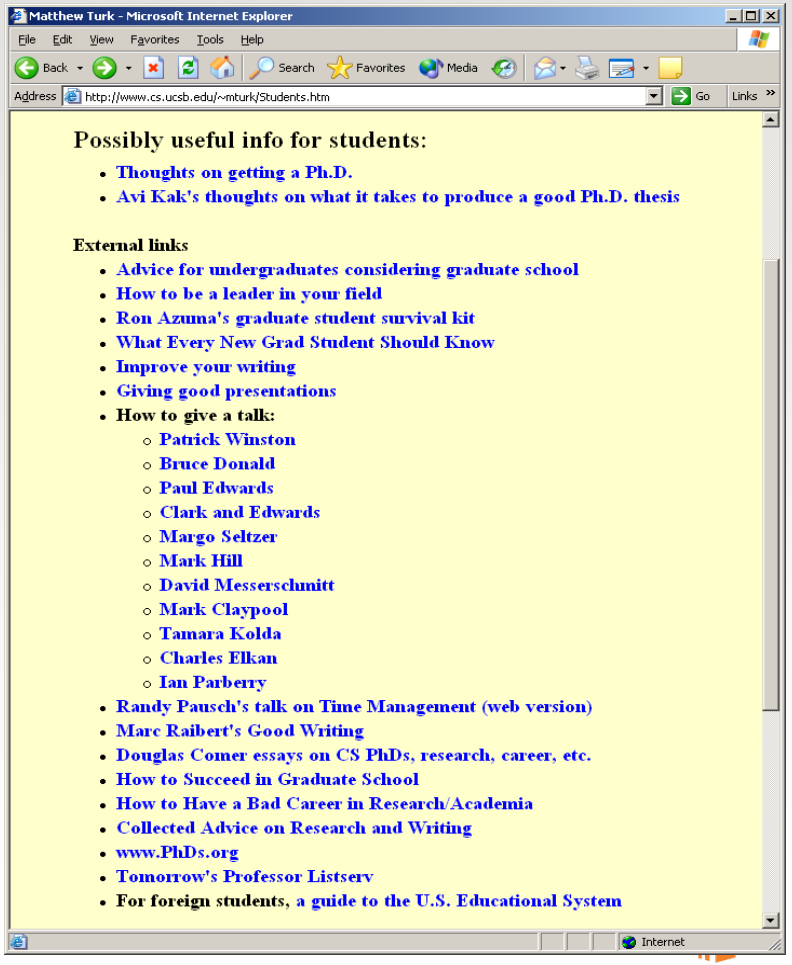
- Work hard, networking, think, read, program, experiment, build, study, practice, ….
  - So little time and so much to do!!

- 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

AUBURN UNIVERSITY
SAMUEL GINN COLLEGE OF ENGINEERING

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 ...

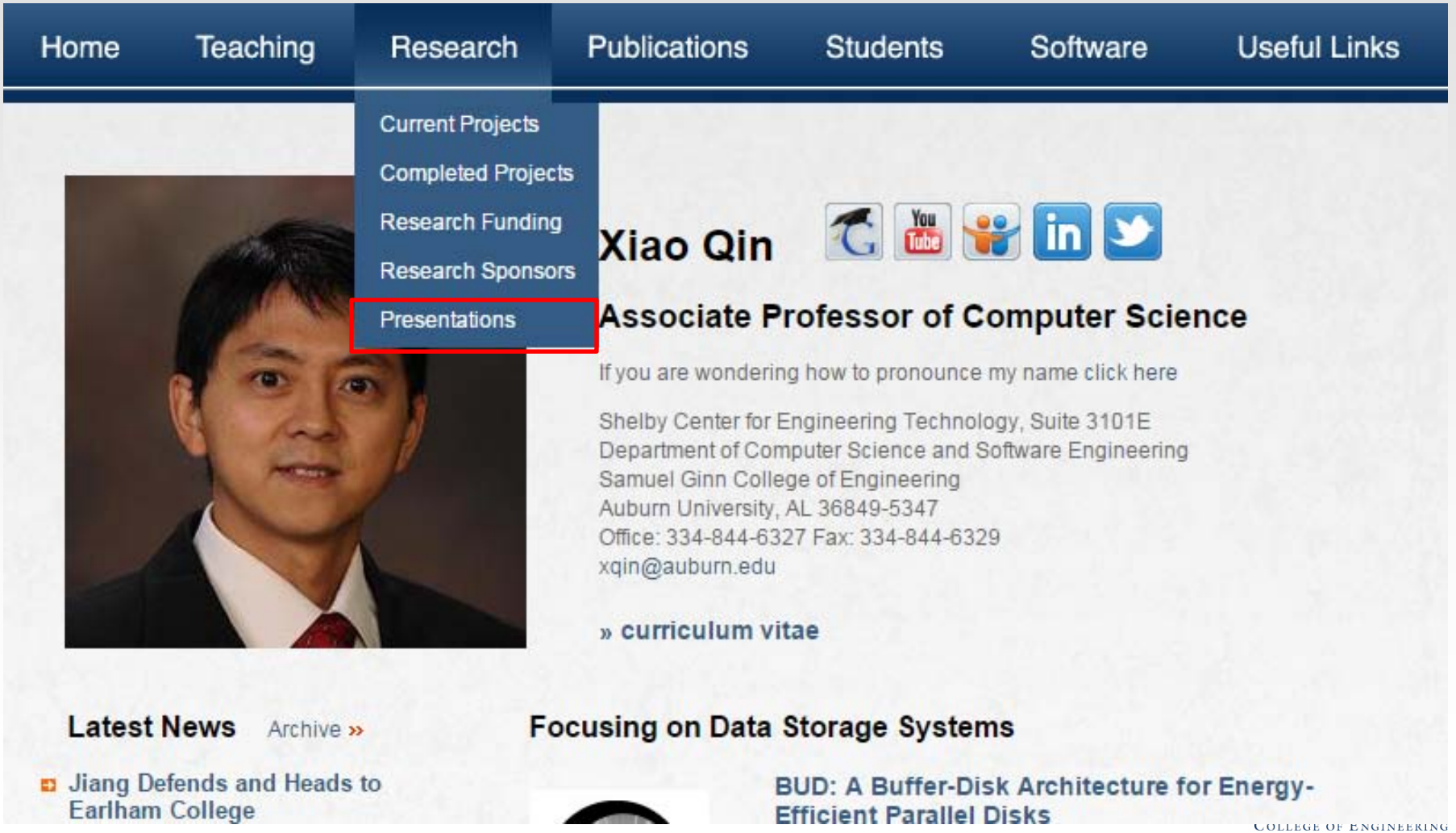
www.eng.auburn.edu/~x**qin**/ - Cached - Similar

AUBURN UNIVERSITY

SAMUEL GINN COLLEGE OF ENGINEERING

My webpage

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



Download Slides at slideshare

http://www.slideshare.net/xqin74

Home | Teaching | Research | Publications | Students | Software | Useful Links

Xiao Qin's Presentations

AUBURN UNIVERSITY

Presentations

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

Download our slides from SLIDESHARE PRESENTATIONS

Download the source code of our data placement module for

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 Int'l Heterogeneity in Computing Workshop, Atlanta, Georgia, April 2010. [PPT]
3. An Application-Oriented Approach for Computer Security Education," Invited Talk at the Information Security and Computer Applications Conference, Feb. 25, 2011. [PPT]
4. Energy-Efficient Prefetching with Buffer Disks for Cluster File Systems. IEEE

slideshare present yourself HOME GO PRO CHANNELS ZIPCAST UPLOAD Search

Connect Xiao Qin

Xiao Qin

Analytics Newsfeed Presentations Documents Videos Following Favorites More

Username xqin74

Post on wall

Private message

Presentations 11
Public 11
Private 0

Tags (23)

application-oriented clusters computer security data placement disk systems education energy efficiency energy efficient going to graduate school hdfs heterogeneous cluster high-performance improve mapreduce papers read real-time parallel applications research papers resource management scheduling secure security-aware write

Presentations

Latest Most viewed

Improve energy efficienc...
2 months ago, 103 views
Edit

Security-Aware Schedulin...
2 months ago, 80 views
Edit

Going to graduate school
2 months ago, 51 views
Edit

Energy efficient resourc...
2 months ago, 96 views
Edit

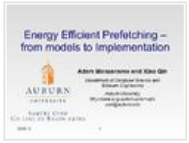
How to read papers
2 months ago, 56 views
Edit

HDFS-HC: A Data Placemen...
2 months ago, 288 views
Edit

How to write papers, pa...
2 months ago, 52 views
Edit

How to write papers, pa...
2 months ago, 65 views
Edit

An Application-Oriented ...
2 months ago, 107 views
Edit



Energy Efficient Prefetc...
11 months ago, 263 views
Edit

How to be a successful u...
12 months ago, 105 views
Edit

Summary

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

- **Download the slides at**

http://www.slideshare.net/xqin74

AUBURN UNIVERSITY
SAMUEL GINN COLLEGE OF ENGINEERING

Questions