How to Leverage Research Universities

Carlos Maltzahn
Center for Research in Open Source Software
UC Santa Cruz

- CROSS
- Engagement
- CROSS Incubator
- Conclusions
Research Universities
Students

Most get a job in industry after their degree
Undergrad & Master students

Go for grades and experience
PhD students

Go for publications & experience
Research publications

Research results for engineers (and VCs)
What about Software?

PhDs in Science & Engineering in the US:
2017: 54,664
2016: 54,862
2015: 54,901
... [NCSES]

Likely that some of them created significant research prototype software
Research Prototype Software Life Cycle

PhD students build amazing research prototype software

Available as OSS as part of publications

Software is left behind when students graduate
Available as OSS as part of publications

Software is left behind when students graduate
Available as OSS as part of publications

Software is left behind when students graduate

Unless PhD students become OSS leaders
PhD students turned OSS Leaders

Sage Weil (Ceph) - UC Santa Cruz
Matei Zaharia (Spark) - UC Berkeley
Chris Lattner (LLVM, Clang) - UIUC
Margo Seltzer (Berkeley DB) - UC Berkeley
...
Efficient Research Delivery

Implement efficient delivery of software = efficient delivery of research

Create and shape new markets

Create significant value and attract network of talent
How to Leverage Research Universities

Carlos Maltzahn
Center for Research in Open Source Software
UC Santa Cruz

Research Universities
CROSS
CROSS Incubator
Engage!
Conclusions
• **Bridges gap between student research & open source projects**
• Funded by Sage Weil endowment & corporate memberships

• Educate the next generation of OSS leadership
• Leverage OSS culture in university research
• Incubate work beyond graduation to reach critical mass
Governance

Industry Advisory Board

Advisory Committee

Stephanie Lieggi, Assistant Director

Carlos Maltzahn, Director

Lavinia Preston, Admin Assistant

Doug Cutting (Chief Architect, Cloudera)
Founder of numerous groundbreaking open-source projects, including Luene, Nutch, Avro, Hadoop

Karen Sandler (Executive Director, Software Freedom Conservancy)
Former executive director of GNOME Foundation, former general counsel of Software Freedom Law Center

Sage Weil (Chief Architect of Ceph, Red Hat)
Creator of Webebling, founder of DreamHost, founder and CTO of Inktank, recipient of O’Reilly Open Source Award

James Davis (Professor of Computer Science and Engineering)
He was Founding Director of UCSC Center for Entrepreneurship, sits on advisory councils for startups and nonprofits. Interested in research applying technology to address global social issues, and has developed an award winning course around this theme.
Advisory Committee

Doug Cutting (Chief Architect, Cloudera)
Founder of numerous groundbreaking open-source projects, including Lucene, Nutch, Avro, Hadoop

Karen Sandler (Executive Director, Software Freedom Conservancy)
Former executive director of GNOME Foundation, former general counsel of Software Freedom Law Center

Sage Weil (Chief Architect of Ceph, Red Hat)
Creator of WebRing, founder of DreamHost, founder and CTO of Inktank, recipient of O’Reilly Open Source Award

James Davis (Professor of Computer Science and Engineering)
He was Founding Director of UCSC Center for Entrepreneurship, sits on advisory councils for startups and nonprofits. Interested in research applying technology to address global social issues, and has developed an award winning course around this theme.
Current Fellowships

3 Research Fellows (see cross.ucsc.edu) and 3 Incubator Fellows (see below)
CROSS Operations

- Modelled after NSF’s I/UCRCs
- Adds open-source software focus
- Sustained through membership fees
How to Leverage Research Universities

Carlos Maltzahn
Center for Research in Open Source Software
UC Santa Cruz

Research Universities
CROSS
CROSS Incubator
Engage!
Conclusions
CROSS Incubator

Currently funding 3 Fellows

Current Fellows
Qualifications
Benefits
Current Fellows

Currently funding 3 Fellows:
Jeff LeFevre, Kate Compton, Ivo Jimenez
Jeff LeFevre: Skyhook

Programmable Storage for Databases

https://users.soe.ucsc.edu/~jlefevre/
Kate Compton: Tracery 2 & Chancery
Tracery but with finite state machines instead of grammars.

Tracery: generate text, graphics and more

Tracery is a super-simple tool and language to generate text, by GalaxyKate. It's been used by middle school students, humanities professors, indie game developers, professional bot makers, and lots of regular people, too. Give it a try today!

GET THE REPO  TRY AN ONLINE TUTORIAL  JUMP INTO THE EDITOR  MAKE A TWITTERBOT
DOWNLOAD AND PRINT A HELPFUL ZINE  READ AN ACADEMIC PAPER

https://www.galaxykate.com/
Ivo Jimenez: Black Swan

The Practical Reproducibility Platform

https://users.soe.ucsc.edu/~ivo/
Incubator Fellow Qualifications

Graduated with a Ph.D. and is a well-published expert

Starts out with a significant code base from the Ph.D. project

Leverages at least one well-established OSS community

Wants to become an OSS leader
Benefits

Postdoc position with salary + benefits

Regular review by Advisory Committee and Industry Advisory Board

**Community seeding** via "Research Experience" Programs

- Tap into pool of students who need projects to work on
- CROSS is Google Summer of Code Mentor Organization

Great community management training

Great driver for community infrastructure
How to Leverage Research Universities

Carlos Maltzahn
Center for Research in Open Source Software
UC Santa Cruz
Engage!

Join university research centers like CROSS

Membership fees encourage OSS strategies & techniques in research and tech transfer

IAB meetings are twice a year (mid March and beginning of October)

Join & contribute to OSS incubator projects (no membership needed)

Participate in 2-day CROSS Research Symposium (next is October 2-3)

Shows off student work at CROSS and other UC Santa Cruz programs
### October 24-25, 2016

#### Monday, October 24 in the Simulatur, E2-180

**8:15-8:45**
Registration and Continental Breakfast, Lower Engineering Courtyard

**8:45-9:00**
Welcome

**9:00-9:30**
Keynote Address: Josh Stuart (UCSC), Professor of Biomolecular Engineering, Associate Director of OBSE for Cancer and Stem Cell Genomics, Baskin Endowed Chair

**9:30-10:00**
Presentation Preview Spotlights and Introduction to Workshops

**10:00-10:30**
BREAK – Refreshments and Discussion, Lower Engineering Courtyard

**10:30-12:00**
**Session I**

**Workshops**

<table>
<thead>
<tr>
<th>Secure, Real-time Sharing of Cancer Gene Information</th>
<th>Data-Driven Dynamic Networked Systems</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chair: Rob Currie, CTO, Genomics Institute Simulatur, E2-180</td>
<td>Chair: Ricardo Sanfelice Engineering 2, STHFloor, Room 599</td>
</tr>
<tr>
<td>- The Cancer Gene Trust (Rob Currie)</td>
<td>- Models, Algorithms, and Evaluation for Autonomous Mobility-On-Demand Systems (Marco Pavone)</td>
</tr>
<tr>
<td>- ADAM (Frank A. Nothelfer)</td>
<td>- Challenges of Implementing Incentive Mechanisms for Reducing Infrastructure Congestion (John Mussaich)</td>
</tr>
<tr>
<td>- Towards the Internets of the Future (Katia Orzacza)</td>
<td>- Towards the Internets of the Future (Katia Orzacza)</td>
</tr>
</tbody>
</table>

**12:00-1:30**
LUNCH – Catered by Fatafel of Santa Cruz

**1:30-3:00**
**Session II**

- Hull (Henne Schmitt)
- UCSC Genomics Core Architecture (Brian O'Connor)
- From Distributed Robotics to Cloud Robotics (Stefano Carpin)
- Scalable Collective Reasoning over Network Data (Luz Getoor)

**3:00-4:00**
BREAK & Poster Session – Refreshments and Discussion, Lower Engineering Courtyard

**4:00-5:00**
**Session III**

**Keynote Address: David Haasler (UCSC), Investigator at Howard Hughes Medical Institute, Distinguished Professor of Biomolecular Engineering, Scientific Director of UC Santa Cruz Genomics Institute, Director of Cancer Genomics Hub, Scientific Co-Director of California Institute for Quantitative Biosciences (QB3)**

**5:00-8:00**
Oktoberfest Dinner – Lower Engineering Courtyard

#### Tuesday, October 25 in the Simulatur, E2-180

**8:15-8:45**
Registration and Continental Breakfast, Lower Engineering Courtyard

**8:45-9:00**
Welcome

**9:00-9:30**
Keynote Address: Karen Sandifer, Executive Director of Software Freedom Conservancy

**9:30-10:00**
Presentation Preview Spotlights and Introduction to Workshops

**10:00-10:30**
BREAK – Refreshments and Discussion, Lower Engineering Courtyard

**10:30-12:00**
**Session I**

**Workshops**

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Chairs: Carlos Matlazhn, Jishen Zhao Simulatur, E2-180</td>
<td>Chairs: Nic Brumwell, John Gustafson Engineering 2, STHFloor, Room 599</td>
</tr>
<tr>
<td>- Zing: distributed shared-log for software-defined storage (Noah Watkins, UCSC)</td>
<td>- Keynote Address: John Gustafson, Visiting Scientist at A&quot;STAR: “Unum systems for better computational arithmetic”</td>
</tr>
<tr>
<td>- Mantle: A Programmable Metadata Load Balancer for the Ceph File System (Michael Sevilla, UCSC)</td>
<td>- General discussion of unums</td>
</tr>
<tr>
<td>- Persistent Memory Workload Characterization: A Hardware Perspective (Bhaskar Jupidi and Xiao Liu, UCSC)</td>
<td></td>
</tr>
</tbody>
</table>

**12:00-1:30**
LUNCH – Catered by Scopazzi’s

**1:30-3:00**
**Session II**

- Keynote Address: Pankaj Mehra, VP & Senior Fellow, Sandisk / Western Digital: “Evolutionary developments, revolutionary implications: persistent memory in the data center”
- Complexity-effective architecture solutions for persistent memory systems (Matheus Ogliari, UCSC)
- Measurements For A Tailored Number System (Andrew Shevemaker, LANL)
- Unum implementation in Julia (Isaac Yonemoto, A"STAR)

**3:00-3:30**
BREAK – Refreshments and Discussion, Lower Engineering Courtyard

**3:30-5:00**
**Session III**

- Elastic Databases for the Cloud (Jeff LeFevre, UCSC)
- Ingo: Ambidextrous Congestion Control in Data Centers (Andrew Shevemaker, LANL)
- The Popper Convention: Making Reproducible Systems Evaluation Practical (Ivo Jimenez, UCSC)
- Measurements For A Tailored Number System (Andrew Shevemaker, LANL)
- LULESH execution with Unums in C/C++ (Scott Lloyd, LLNL)
- Progress report of the CROSS Unum project (Andrew Kofias, UCSC)

---

30.
Coming up: October 3-4, 2018

Keynote Speakers:

Cat Allman (Google), Anh-Anh Doan (UW Madison), Jay Kreps (Confluent)

Wednesday, October 3

Registration and Continental Breakfast, Lower Engineering Courtyard

PLENARY - Simmons Hall, E2-180

8:45-9:00
Welcome by Carla Mahlman, CROSS Director

9:00-10:00
Plenary Talk: Open Source Software for Provenance for Data Integration - Experience and Lessons Learned

Anh-Anh Doan, University of Wisconsin-Madison

10:15-10:35
Presentation of the Keynotes and Introduction to Workshops (Chair: Stephanie Lagg)

10:35-10:45
BREAK - Refreshments and Discussion, Lower Engineering Courtyard

WORKSHOPS

Data Cleaning

Chief, Data Steward, Simmons Hall, E2-180

Faizan Qureshi (IBM Research), Exploring Knowledge Graphs via Natural Language Interfaces

Berkay Barlas (Stanford University), Codes, Crunching, and the Berlin-Warsaw Alliance in the Fight for Clean Data

10:45-11:00 Section 1

Data Cleaning

Chief, Data Steward, Simmons Hall, E2-180

Architecture Specification

Chair: John Sorensen, Engineering 1, 5th Floor, Room 599

Jack Koning, IBM, Directing the Level of Abstraction in Digital Design with Cadence and PARLIS

11:00-11:15 Section 2

Los Angeles (UCLA), Algorithmic Aspects of Data-Driven Database Repairing

Ryan Zink (UCLA), A Simple Approach to Consistent Query Answering

11:15-11:30 Section 3

Data Cleaning

Chief, Data Steward, Simmons Hall, E2-180

Architecture Specification

Chair: John Sorensen, Engineering 1, 5th Floor, Room 599

11:30-11:45 Section 4

11:45-12:00 Section 5

12:00-1:00 Lunch - Lower Engineering Courtyard

1:00-1:15
Karniel Druckman, Behemoth of Code: 14 years of work

Cat Allman, Google

1:15-1:30 Section 6

Data Cleaning

Chief, Data Steward, Simmons Hall, E2-180

Architecture Specification

Chair: John Sorensen, Engineering 1, 5th Floor, Room 599

1:30-1:45 Section 7

1:45-2:00 Section 8

2:00-2:15 Section 9

2:15-2:30 Section 10

2:30-2:45 Section 11

2:45-3:00 Section 12

3:00-3:15 Section 13

3:15-3:30 Section 14

3:30-3:45 Section 15

3:45-3:55 Section 16

4:00-4:15 Section 17

4:15-4:30 Section 18

4:30-4:45 Section 19

4:45-5:00 Section 20

5:00-5:30 Poster Session - Refreshments and Discussion, Lower Engineering Courtyard

5:30-7:30 CAFETERIA Buffet - Lower Engineering Courtyard
How to Leverage Research Universities

Carlos Maltzahn
Center for Research in Open Source Software
UC Santa Cruz

- CROSS
- Engage!
- Research Universities
- CROSS Incubator
- Conclusions
Conclusions

Traditionally research universities produced students and papers

Now there is a 3rd "product": OSS leadership

CROSS is creating structures which enable PhD students to turn into OSS leaders

Universities with their large student pool spanning undergraduate and graduate programs are ideal starting points for creating communities

Contact: cross-info@ucsc.edu, web: cross.ucsc.edu, youtube: bit.ly/uccross
How to Leverage Research Universities

Carlos Maltzahn
Center for Research in Open Source Software
UC Santa Cruz