



# Undergrad & Master students

Go for grades and experience



# PhD students

Go for **publications** & experience







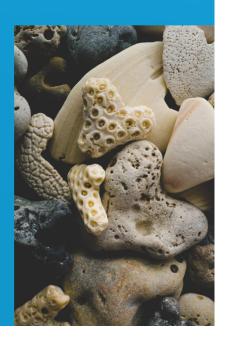
# Research Prototype Software Life Cycle

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



#### PhD students turned OSS Leaders

Sage Weil (Ceph) - UC Santa Cruz

n ceph

Matei Zaharia (Spark) - UC Berkeley

Chris Lattner (LLVM, Clang) - UIUC

Margo Seltzer (Berkeley DB) - UC Berkeley





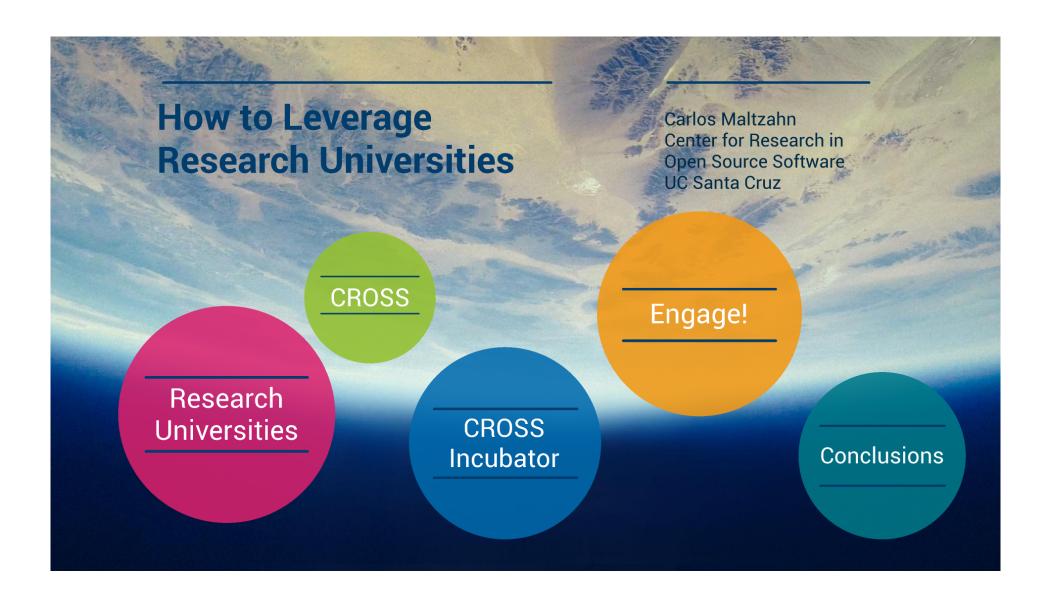
## **Efficient Research Delivery**

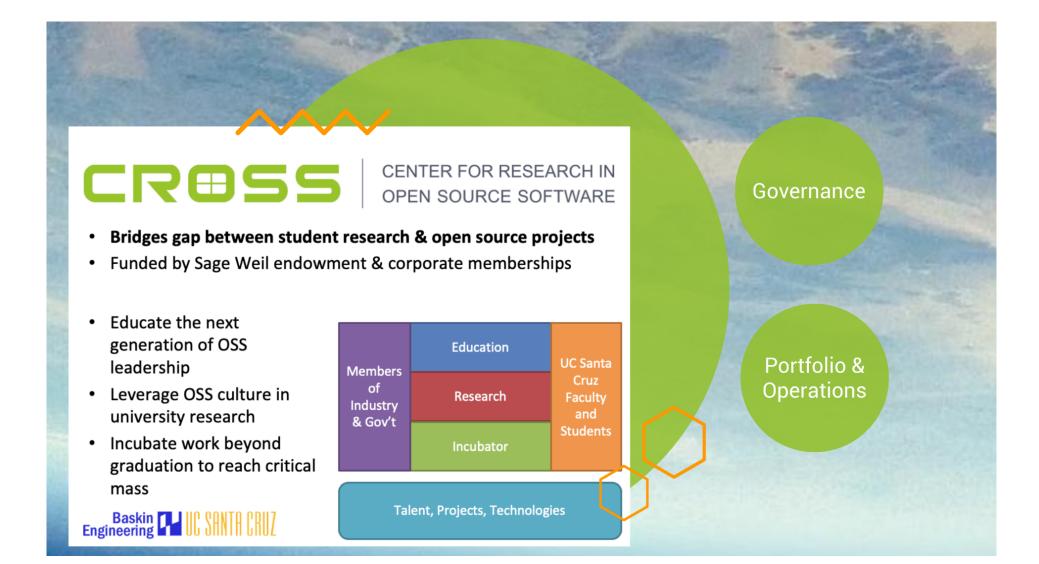
Implement efficient delivery of software = efficient delivery of research

Create and shape new markets

Create significant value and attract network of talent







#### Governance



Stephanie Lieggi, **Assistant Director** 



Director



Lavinia Preston, Admin Assistant

#### **Industry Advisory Board**













#### **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.

# **Advisory Committee**



**Doug Cutting** (Chief Architect, Cloudera)
Founder of numerous groundbreaking open-source projects, including Lucene, <u>Nutch</u>,
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.

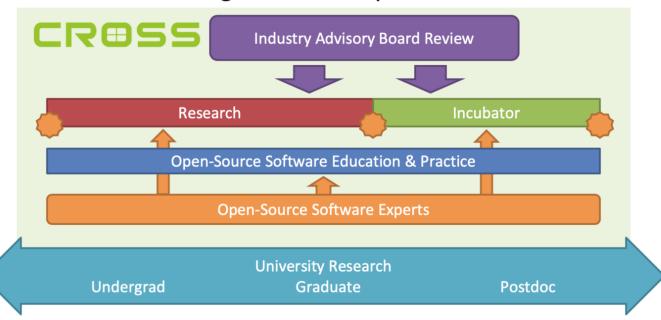
## **Current Fellowships**

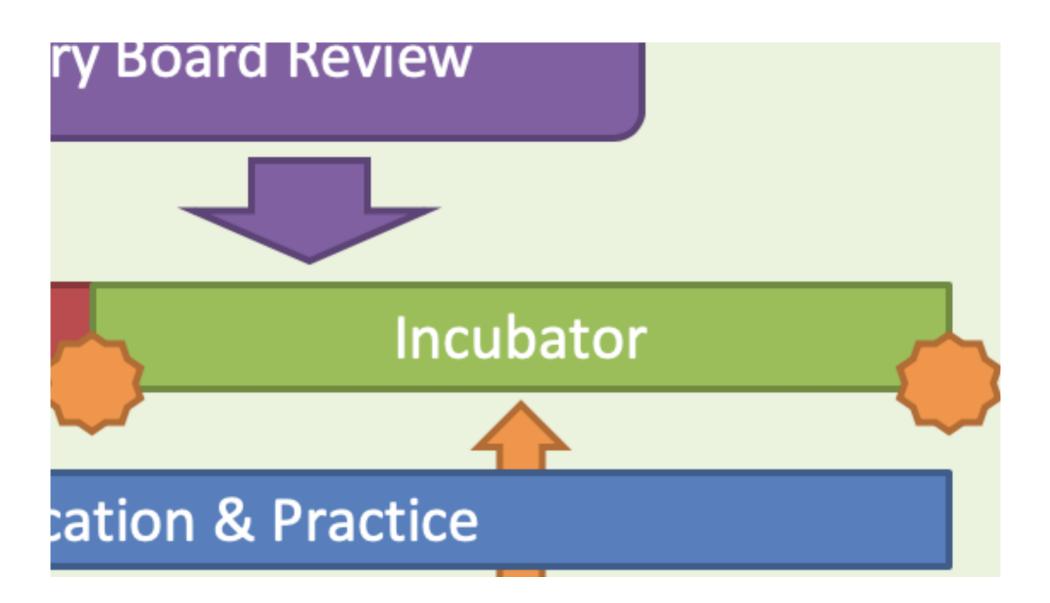
3 Research Fellows (see cross.ucsc.edu) and 3 Incubator Fellows (see below)

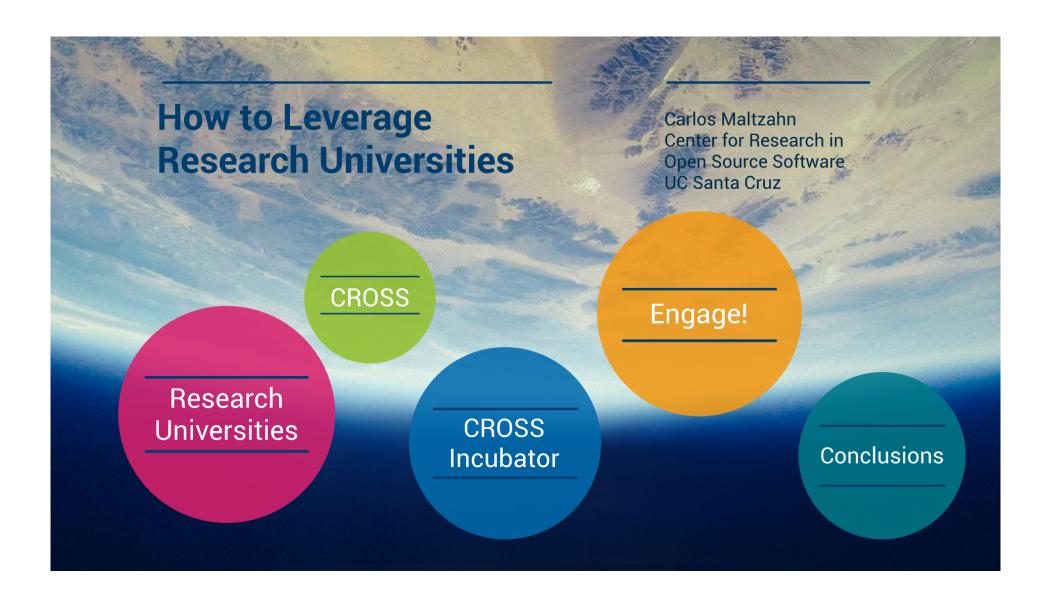


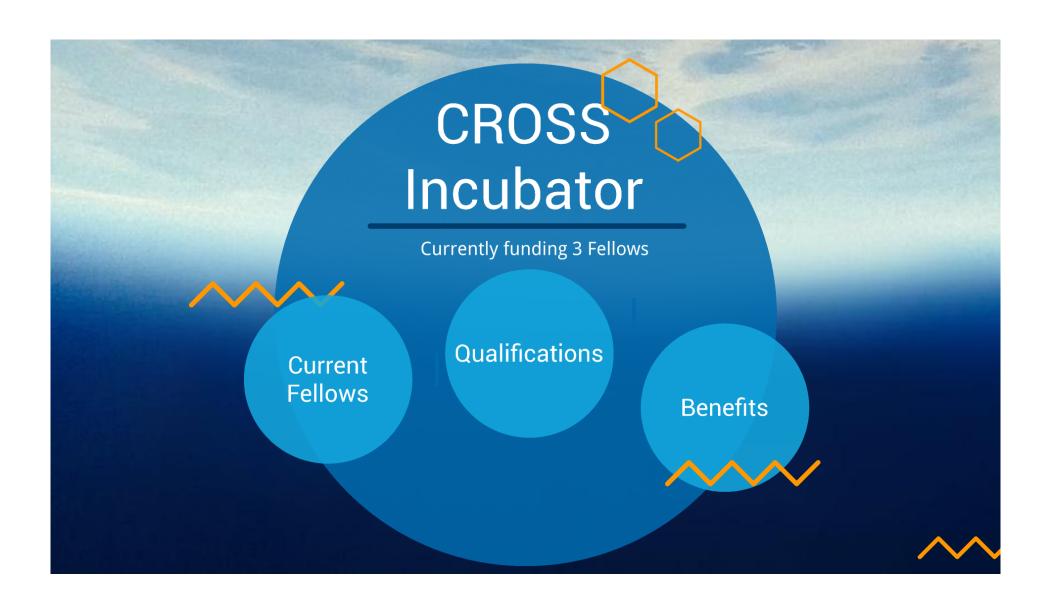
## **CROSS Operations**

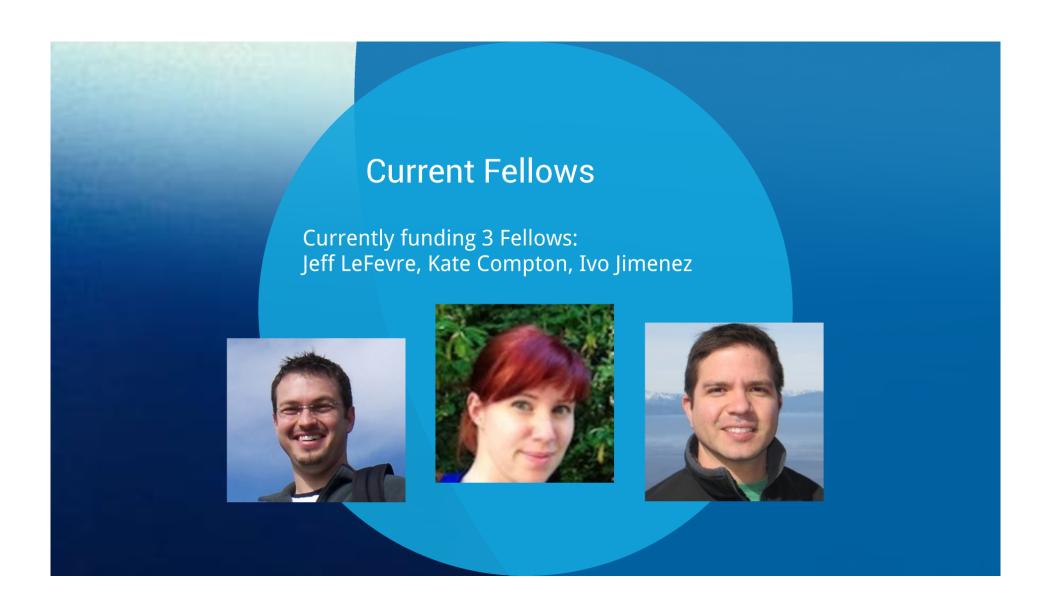
- Modelled after NSF's I/UCRCs
- Adds open-source software focus
- Sustained through membership fees





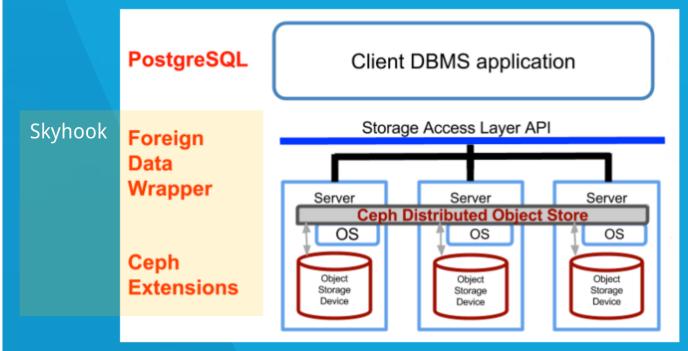






## Jeff LeFevre: Skyhook

**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 <u>GalaxyKate</u>. 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

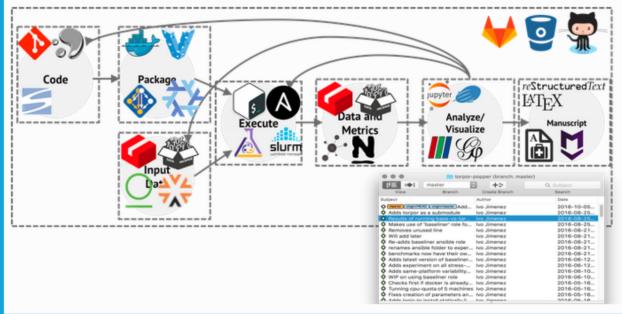
https://www.galaxykate.com/

#### Ivo Jimenez: Black Swan

The Practical Reproducibility Platform







https://users.soe.ucsc.edu/~ivo/

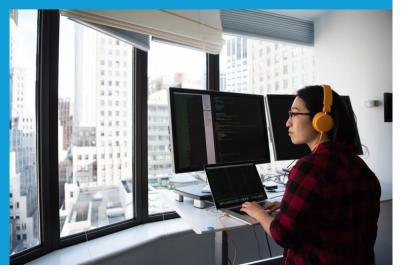
### **Incubator Fellow Qualifications**

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

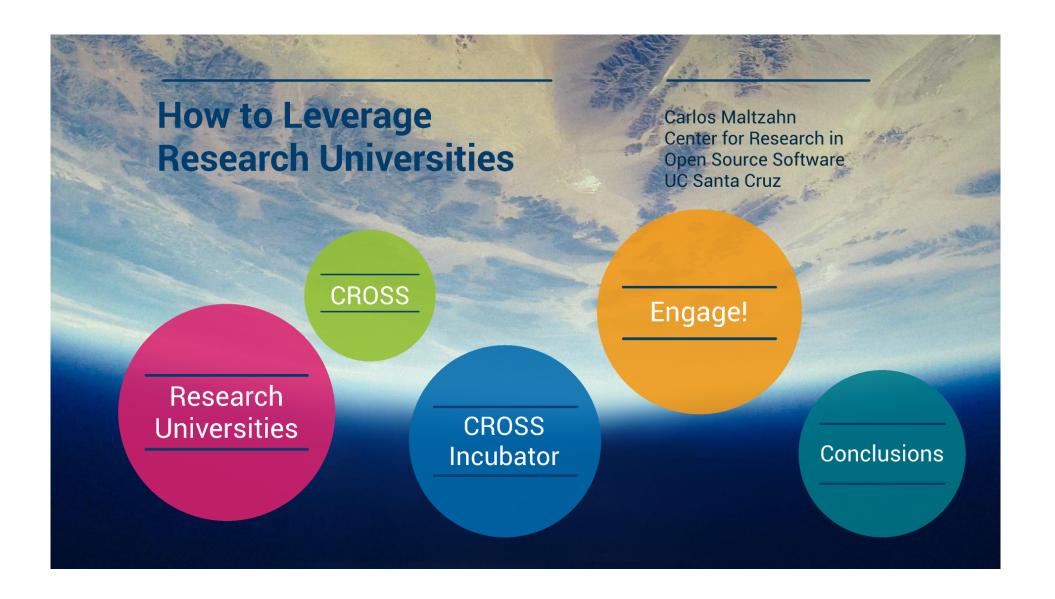
Tap into pool of students who need projects to work on

CROSS is Google Summer of Code Mentor Organization

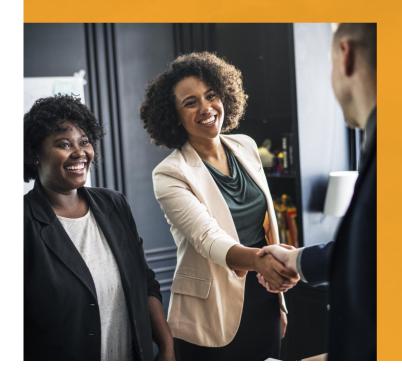
Great community management training

Great driver for community infrastructure





## **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 Simularium, E2-180				
8:15-8:45	Registration and Continental Breakfast, Lower Engineering Courtyard			
8:45-9:00	Welcome			
9:00-9:30	Kevnote Address: Josh Stuart (UCSC).  Professor of Biomolecular Engineering, Associate Director of CBSE 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			
	WORKSHOPS			
	Secure, Real-time Sharing of Cancer Gene Information Chair: Rob Currie, CTO, Genomics Institute Simularium, E2-180	Data-Driven Dynamic Networked Systems  Chair: Ricardo Sanfelice Engineering 2, 5THFloor, Room 599		
10:30- 12:00 Session I	- The Cancer Gene Trust (Rob Currie) - ADAM (Frank A. Nothaft)	- Models, Algorithms, and Evaluation for Autonomous Mobility-On-Demand Systems (Marco Pavone)  - Challenges of Implementing Incentive Mechanisms for Reducing Infrastructure Congestion (John Mussachio)  - Towards the Internets of the Future (Katia Obraczka)		
12:00-1:30				
1:30-3:00 Session II	Toil (Hannes Schmidt)     UCSC Genomics Core Architecure (Brian O'Connor)     Smart Storage in Genomics (Carlos Maltzahn)	- From Distributed Robotics to Cloud Robotics (Stefano Carpin) - Scalable Collective Reasoning over Network Data (Lise Getoor)		
3:00-4:00	BREAK & Poster Session – Refreshments and Discussion, Lower Engineering Courtyard			
4:00-5:00	Keynote Address: David Haussler (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 Simularium, 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 Sandler,			
0:20-10:00	Executive Director of Software Freedom Conserva			
	9:30-10:00 Presentation Preview Spotlights and Introduction to Workshops			
10:00-10:30	00-10:30 BREAK – Refreshments and Discussion, Lower Engineering Courtyard			
	WORKSHOPS			
	Programmability and Architecture of Storage Systems	Towards a Better Computer Arithmetic: A unum (Universal Number) Workshop		
	Chairs: Carlos Maltzahn, Jishen Zhao Simularium, E2-180	Chairs: Nic Brummell, John Gustafson Engineering 2, 5THFloor, Room 599		
10:30-12:00 Session I	- Zlog: distributed shared-log for software-defined storage (Noah Watkins, UCSC) - Mantle: A Programmable Metadata Load Balancer for the Ceph File System (Michael Sevilla, UCSC) - Persistent Memory Workload Characterization: A Hardware Perspective (Bhaskar Jupudi and Xiao Liu, UCSC)	Keynote Address: John Gustafson, Visiting Scientist at A*STAR: "Unum systems for better computational arithmetic"     General discussion of unums		
12:00-1:30				
developments, revolutionary implications: (Andrew Shewmaker, LANL)		- Unum implementation in Julia (Isaac Yonemoto,		
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) - Inigo: Ambidextrous Congestion Control in Data Centers (Andrew Shewmaker, LANL) - The Popper Convention: Making Reproducible Systems Evaluation Practical (Ivo Jimenez, UCSC)	- LULESH execution with Unums in C/C++ (Scott Lloyd, LLNL) - Progress report of the CROSS Unum project (Andrew Klofas, UCSC)		

## October 3-4, 2017

Wednesday, October 4					
8:00-8:45	Registration and Continental Breakfast, Lower Engi	ineering Courtyard			
	PLENARY - Simularium, E2-180				
8:45-9:00	Welcome by Alexander Wolf, Dean, Baskin School of Engineering and by Carlos Maltzahn, CROSS Director				
	ndustry perspective				
9:00-10:00	Chair: Doug Cutting (Cloudera) Panelists: Michell Casbon (Qordoba), Tushar Shanbhag (Linkedin), Raymie Stata (Altiscale), Carlos Maltzahn (UCSC)				
10:00-10:30	Presentation Preview Spotlights and Introduction to	o Workshops (Chair: Stephanie Lieggi)			
10:30-10:50	BREAK – Refreshments and Discussion, Lower En	gineering Courtyard			
	WORKSHOPS	S			
	Computational Genomics	Security in Network Systems			
	Chair: Rob Currie, CTO, Genomics Institute Simularium, E2-180	Chair: Ricardo Sanfelice Engineering 2, 5th Floor, Room 599			
10:50-12:00 Session I	- Toiling Away in Bioinformatics: Toil and Data Workflows in Bioinformatics (Lon Blauvelt, Jesse Brennan)	- CapNet: Security and Least Authority in a Capability- Enabled Cloud (Anton Burtsev) - The Importance of Interpretability in Cyber Security Analytics (Juston Moore) - A Control-Theoretic and Data-Driven Approach to Securing Cyber-Physical Systems and Networks (Fabio			
		Pasqualetti)			
12:00-1:00	12:00-1:00 LUNCH – Lower Engineering Courtyard (Catered by Falafel of Santa Cruz)				
PLENARY - Simularium, E2-180					
1:00-1:45	Keynote: Nithya Ruff - How Companies are Engaged in Open Source				
	Senior Director, Open Source Practice, Comcast				
	WORKSHOPS				
	Reproducibility	Network Systems Applications			
	Chair: Ivo Jiminez (UCSC) Simularium, E2-180	Chair: Ricardo Sanfelice Engineering 2, 5th Floor, Room 599			
1:50-3:10 Session II	- Reproducibility Challenges and Opportunities in	- Controlling Biology: Towards Optimal Design of			
000000000000000000000000000000000000000	Computational Science (Carl Boettiger) - BugSwarm: A Large-Scale Repository of	Synthetic Networks (Marcella Gomez) - Convex Relaxations for Power System State			
	Reproducible Defects, Tests, and Patches (Cindy	Estimation (Yu Zhang)			
	Rubio Gonzalez)				
3:10-3:40	BREAK – Refreshments and Discussion, Lower Engineering Courtyard				
	Open Data Management	OSS-based Energy Systems			
3:40-5:00	Chair: Ivo Jiminez (UCSC) Simularium, E2-180	Chair: Nick Adams Engineering 2, 5th Floor, Room 599			
Session II	- The Democratization of Data for Infrastructure	- An Intelligent Microgrid Energy Management System			
	Management (Selene Arrazolo)	powered by SEADS (Ali Adabi)			
	- Clinical and Computational Genomics in Open	- Solar PV system energy modeling tools and data			
	Science (Rob Currie) (Jessica Forbess)				
5:10-6:00	Poster Session 1 – Refreshments and Discussion, Lower Engineering Courtyard				
6:00-8:00	OKTOBERFEST Barbeque – Lower Engineering Co	purtyard			

	Thursday, Octob	per 5	
8:15-8:45	Registration and Continental Breakfast, Lower Engineering Courtyard		
	PLENARY - Simulariu	m, E2-180	
8:45-9:00	Welcome by Carlos Maltzahn, CROSS Director		
9:00-9:30	Keynote: Allen Samuels - The Consequences of Infinite Storage Bandwidth		
9:00-9:30	Engineering Fellow, Western Digital Corporation		
9:30-10:00	Presentation Preview Spotlights and Introduction	n to Workshops (Chair: Stephanie Lieggi)	
10:00-10:30	BREAK – Refreshments and Discussion, Lower Engineering Courtyard		
	WOR	KSHOPS	
	Programmable Storage	Probabilistic Programming Languages	
	Chair: Carlos Maltzahn Simularium, E2-180	Chair: Claudia Wehrhahn Engineering 2, 5THFloor, Room 599	
10:30-12:00 Session I	DeciStore: Layering is for the Faint of Heart (Noah Watkins)     Malacology: A programmable storage system (Michael Sevilla)     Skyhook/DB: leveraging programmable object storage toward database elasticity in the cloud (Jeff LeFevre)	- Flexible programming of hierarchical modeling algorithms (MCMC and more) and compilation of R using NIMBLE (Christopher Paciorek) - Probabilistic Soft Logic: A Scalable Probabilistic Programming Language for Modeling Richly Structured Domains (Golnoosh Farnadi)	
12:00-1:30	LUNCH – Lower Engineering Courtyard (Catered by Scopazzi's) & Poster Session 2		
	Architecture of Storage Systems	Applications of Bayesian Statistical Learning	
	Chair: Heiner Litz Simularium, E2-180	Chair: Abel Rodriguez Engineering 2, 5THFloor, Room 599	
1:30-2:45 Session II	- Memshare: a Multi-tenant Key-value Cache (Asaf Cidon) - Persistent Memory: Opportunities and Challenges (Niket Agarwal) - Key value SSD: a scalable smart storage for object storage (Yang Seok Ki)	- Distributed Spatial Kriging: Scalable Bayesian	
2:45-3:15	BREAK – Refreshments and Discussion, Lower Engineering Courtyard		
	Architecture of Storage Systems	Unum: Better Computer Arithmetic	
3:15-4:30	Chair: Yishen Zhao Simularium, E2-180	Chair: John Gustafson Engineering 2, 5THFloor, Room 599	
Session III	Write-Optimized DRAM+NVM Hybrid Memory Management (Xiao Liu)     Efficient Hardware-based Undo+Redo Logging for Persistent Memory Systems (Matheus Ogleari)	Beating Floating Point at its Own Game: Posit Arithmetic (John Gustafson)     Demonstration of the Fused Dot Product in Posit Environments (Isaac Yonemoto)	

Coming up: October 3-4, 2018

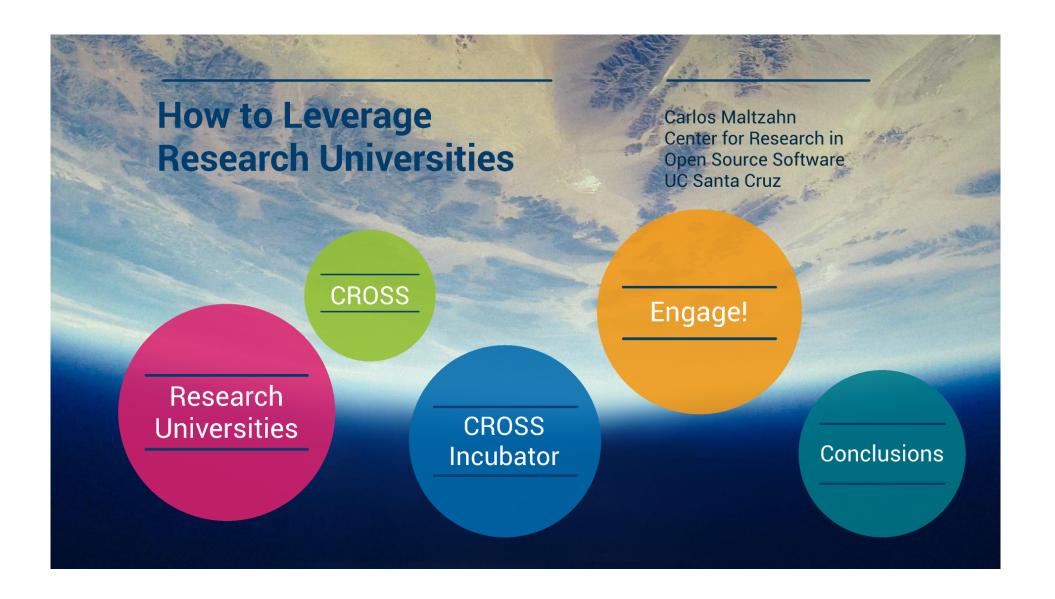
#### Keynote Speakers:



Cat Allman (Google	, AnHai Doan (	U Wisc Madison),	Jay Kreps	(Confluent
--------------------	----------------	------------------	-----------	------------

	Wednesday, Octo	ober 3		
8:00-8:45	Registration and Continental Breakfast, Lower Engir	neering Courtyard		
	PLENARY - Simularium, E2-180			
8:45-9:00	Welcome by Carlos Maltzahn, CROSS Director			
9:00-10:00	Keynote.  Developing Open Source Software in Academia for Data Integration - Experience and Lessons Learned Anhal Doan, University of Wisconsin-Madison			
10:00-10:25		Workshops (Chair: Stephanie Lieggi)		
10:25-10:45	BREAK - Refreshments and Discussion, Lower Eng			
	WORKSHOP			
	Data Cleaning	Architecture Specification		
10:45-12:00 Session I	Chair: Phokion Kolaitis Simularium, E2-180	Chair: Jose Renau Engineering 2, 5th Floor, Room 599		
	Fatma Ozcan (IBM Research), Exploring Knowledge Graphs via Natural Language Interfaces	Jack Koening (SiFive), Raising the level of abstraction in digital design with Chisel3 and FIRRTL		
	Frederic Sala (Stanford University), Codes, Embeddings, &Non-Euclidean Geometry: Unexpected Allies in the Fight to Clean Data	Georgios Michelogiannakis (Lawrence Berkeley NL), Open Hardware: How Open Source Designs Will Drive the Next Generation of HPC Systems		
		Chris Lavin (Xilinx), RapidWright: Enabling Custom Crafted FPGA Implementations via an Open Source Framework		
12:00-1:00	LUNCH – Lower Engineering Courtyard			
1:00-1:45	Keynots: Geogle Summer of Cede: 14 years of trust Cat Allman, Google			
	Data Cleaning	Architecture Specification		
	Chair: Phokion Kolaitis Simularium, E2-180	Chair: Jose Renau Engineering 2, 5th Floor, Room 599		
1:55-3:10 Session II	Ester Livshits (Technion), Algorithmic Aspects of Least-Effort Database Repairing	Patrick Lysaght (Xilinx), PYNQ, Python Productivity for Zynq		
	Akhil Dixit (UCSC), SAT Solvers - A Promising	Rafael Trapani Posignolo (UCSC), Igraph		
	Approach to Consistent Query Answering	Panel Discussion: Open Source Hardware Main Challenges		
3:10-3:40	BREAK - Refreshments and Discussion, Lower En	gineering Courtyard		
	Genomics & Privacy	Practical Reproducibility		
3:40-4:55 Session III	Chair: Benedict Paten Simularium, E2-180	Chair: Ivo Jimenez Engineering 2, 5th Floor, Room 599		
	Benedict Paten (UCSC), Genomics & Privacy	Ivo Jimenez (UCSC), Black Swan: The Popper Reproducibility Platform		
	Russell Corbett-Detig (UCSC), The Past and Future of Population Genomics	Rolando Garcia (UC Berkeley), Context: The Missing Piece in the Machine Learning Lifecycle		
	Abhradeep Guha Thakurta (UCSC), Distributed Private Machine Learning	Chris Holdgraf (Berkeley Institute for Data Science), Open tools and services for practical reproducibility in		
	Robert Currie (UCSC), Cross Border Lambdas in	the cloud		
	Pediatric and Hereditary Cancer			
4:55-5:30 5:30-7:30		wer Engineering Courtyard		

	Thursday, Octo	ber 4		
8:15-9:00	Registration and Continental Breakfast, Lower Engineering Courtyard			
	PLENARY - Simulariu	m, E2-180		
9:00-9:15	Welcome by Carlos Maltzahn, CROSS Director			
9:15-10:00	Keynote: Introduction to Streaming Data and Stream Processing with Apache Kafka Jay Kreps. Confluent			
10:00-10:25	Presentation Preview Spotlights and Introduction to	Workshops (Chair: Stephanie Lieggi)		
10:25-10:50	BREAK - Refreshments and Discussion, Lower Eng	, , , , , , , , , , , , , , , , , , , ,		
10.20-10.00	WORKSHOPS			
	Eusocial Storage: Shades of the InStore Compute Spectrum	Bayesian Statistical Learning		
	Chair: Philip Kufeldt Simularium, E2-180	Chair: Claudia Wehrhahn Engineering 2, 5thFloor, Room 599		
10:50-12:15 Session I	Philip Kufeldt (UCSC), Eusocial Storage Intro  Yang Seok Ki (Samsung, InStore Compute: Preset Functions	Christopher Paciorek (UC Berkeley), Probabilistic programming in practice: Using, programming, and sharing hierarchical modeling algorithms such as MCMC and Sequential Monte Carlo using NIMBLE		
	Noah Watkins (Red Hat)& Jeff Levre (UCSC), Instore Compute: Interpreted Functions	Claudia Wehrhahn (UCSC), Variable selection for assessing red wine quality: a NIMBLE solution		
	Scott Shadley (NGD Systems) Instore Compute: General Programming			
12:15-1:30	LUNCH - Lower Engineering Courtyard			
	Storage Systems	Smart Power and Cyber-physical Systems		
	Chair: Carlos Maltzahn Simularium, E2-180	Chair: Yu Zhang Engineering 2, 5th Floor, Room 599		
1:30-2:45 Session II	Noah Watkins (Red Hat), Programmable Storage Michael Sevilla (TidalScale), Scalable Global Namespaces	Ming Jin (UC Berkeley), Stability-certified Smooth Reinforcement Learning: A Control-theoretic Perspectiv Yihsu Chen (UCSC), A Power Market Model in Presence of Strategic Prosumers		
	Yiming Zhang (UCSC), Study of QoS in Ceph	Ram Rajagopal (Stanford), Online Learning and Optimization in Distributed Energy System		
2:45-3:15	BREAK – Refreshments and Discussion, Lower Eng	gineering Courtyard		
	Storage Systems	Smart Power and Cyber-physical Systems		
3:15-4:30 Session III	Chair: Carlos Maltzahn Simularium, E2-180	Chair: Yu Zhang Engineering 2, 5th Floor, Room 599		
	Timothy Feldman (Seagate), The Edge of Storage Reza NasiriGerdeh (UCSC): Exploring expressive data	Ruojia Xi (UC Berkeley), Privacy Preservation in Cyber Physical Systems		
	access policies in data-intensive science domains  Jeff LeFevre (UCSC), Skyhook: leveraging	Atif Maqsood (UCSC), Optimal Rotational Load Shedding via Bilinear Integer Programming		
	programmable object storage toward database elasticity	Hongcai Zhang (UC Berkeley), Data-driven Approach t Hamessing Flexibility of Large-scale Distributed Energy Resources for Power Grids		



#### 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



CENTER FOR RESEARCH IN OPEN SOURCE SOFTWARE

