# UCSC's Perspective on File System and I/O Research and Education

Scott A. Brandt
Computer Science Department
University of California, Santa Cruz


# What Gary Asked Us To Do

- Current curricula for FSIO
- Center thrusts
- Issues the HEC community needs to address to build a stronger pipeline of people in this area
  - Especially for the US, given that HEC is part of the US Competitiveness program
- Wider discussion: "Rising above the Gathering Storm"
  - US college students are not entering technical fields
  - CS role models: college drop-out billionaires
  - The top 20% of CS high school students are going to gamer developer conferences, etc.


# Summary

- Too few domestic students are entering Science and Engineering
  - Especially women and minorities
- Those that do tend to go to industry
  - And industry focuses on development, not research
- Even fewer see high-end computing, I/O, and storage as a career path


# Why? (I)

- Not exciting
  - CS seen as the difficult career path best suited for asocial loners
  - Too focused on the tools/process
  - Not seen to have real impact on society
- Jackpot mentality
- Lower job security leads to a focus on salary
  - People expect to change jobs frequently
  - Programming jobs expected to be outsourced
  - Lower salaries make academic and lab careers less inviting


# Why? (II)

- Lower education/research funding makes academic careers less inviting
  - And dropping enrollments ⇒ fewer positions
  - And grad funding harder to obtain
  - Do I want to do what my advisor does? No!
  - Especially true of HEC, I/O, and storage
- HEC and I/O not viewed as first-class sub-disciplines of CS despite size/importance of industry
  - Not visible/valued ("Plumbing isn't exciting")
  - Not viewed as interesting or broad enough
  - Lack critical mass ("One 'device' in this room would take out most of the storage community")
- "Nobody makes money on HEC"


# What We Need To Do

- Attract
  - Change image of science, CS, HEC, and I/O
  - Create exciting opportunities
  - Increase visibility of successes, results, applications
- Train
  - Create new programs, courses, paths
  - Increase focus on HEC, I/O, storage
- Place
  - Create paths that lead to attractive careers
- Retain
  - Create new opportunities
  - Increase availability of research funding
  - (ahem) Improve salaries


# UCSC Computer Science Department

- Largest grad program on UCSC campus (~140 students)
  - Currently offer BA, BS, MS, and Ph.D.
  - Introducing BS/MS program
- Centers of Excellence: Focused research groups targeting specific research areas
  - Storage (3+1)
  - Graphics/Scientific Visualization (3)
  - Machine Learning (2)
  - Databases (3)
  - Bioinformatics (0 -- Now its own department)
- All grad students participate in research (and some ugrads)
  - Grad courses are very research, project-intensive
  - Master's Project, Master's Thesis, or Ph.D. Thesis
- New Program: Interactive Game Engineering


# UCSC Storage Systems Research Center (SSRC)

- 4 faculty (3 regular, 1 research) + affiliates
- ~20 students (grads + undergrads)
- Broad research in storage
  - High-performance, QoS, new technologies, rich metadata, object-based storage, archiving, security, prediction, …
- Strong relationship with industry & National Labs
  - Joint research, internships, placement
  - IBM, Symantec, NetApp, HP, Intel, Microsoft, …
  - LANL, Livermore, Sandia
- Targeted courses
  - Undergrad: operating systems, systems programming, storage systems
  - Grad: operating systems, distributed systems, storage systems, and special topics courses
- Open sourcing Ceph object-based storage system
  - Reference implementation, research platform, punching bag, …


# UCSC/LANL Institute for Scalable Scientific Data Management (ISSDM)

- Collaboration with Los Alamos National Laboratory
- Focused on training, recruiting, and retaining top scientists
- Areas
  - Real-time data collection
  - High-performance peta-scale data management
  - Peta-scale information management
  - Others: Machine learning, Scientific visualization, …
- Graduate education
  - Targeted courses
  - Graduate Student Researchers
  - MS CS (Scientific Data Management)
- Outreach
  - Open sourcing software
  - Seminar series
- Distance education
  - Transmitted to LANL & offered by LANL employees


# Parallel Data Storage Institute (PDSI)*

- Broad collaboration between academia and National Laboratories focused on HEC I/O
  - CMU, UCSC, UMich, LANL, SNL, ORNL, PNNL, NERSC
- Goals: Outreach, Application performance, Storage dependability, Protocol/API extensions, Novel mechanisms, Storage automation
- Approach: Collaborative research, workshops, educational initiatives, affiliates

**might* be funded by SCIDAC


# Ideas: Attract

- Get past nerd (plumber?) image
- Break free from OS community
  - Our own conferences, awards, groups, funding opportunities, etc.
- Focus on applications
  - Tell people why it is important & exciting
- Celebrate successes
  - Publicize successes
  - Create awards to recognize people, projects
- Tout size/growth/importance of industry
- Get greater visibility
  - Grand Challenge?


# Ideas: Train

- Increase focus on education, outreach
- Increase targeted funding
  - Dollars are the votes that count most
- Get industry involved
  - PDL and SSRC show that industry cares
  - Deans listen to industry leaders
- Create focused programs, degrees
- Talk about storage in more classes
- Work on Engineering Deans


# Ideas: Place

- Create focused opportunities and paths
- Fellowships, internships, …
- Be visible
- Interact more with academia

# Ideas: Retain

- Create ongoing opportunities, initiatives, …
- Build collaborations
  - Don’t let people disappear into industry
- Make it a good community for its members
  - Celebrate successes!


# Overall

- Keep pushing!