

Viewpoint

How to Build a Bad Research Center

Sharing lessons learned from experiences creating successful multidisciplinary research centers.

A MAJOR CENTER JUST COMPLETED,¹⁰ so I finally have time to collect my thoughts on centers. I have been part of a dozen centers in computer systems, often as director (see the accompanying table). By center, I mean a research project with at least three faculty, a dozen students, and a common vision. This Viewpoint is from the perspective of an academic in computer systems, but I hope it has wider applicability, even beyond academia. I do not advocate centers for all research; a lone researcher is best for many topics. Why care about the Berkeley experience? Alas, establishing credentials is a lot like bragging so let me apologize in advance, which I will also need to do later. The *U.S. News and World Report* ranked universities for the computer systems field four times since 2002. In every survey, our peers rank UC Berkeley first. In addition, the National Research Council published a study of information technology research that led to multibillion-dollar industries.⁶ UC Berkeley was associated with seven such industries, more than any other university, primarily for its system projects.

The Eight Commandments for a Bad Center

Following the template of my earlier pieces^{a,b} I offer Eight Command-

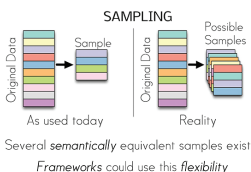
a D. Patterson, <http://www.cs.berkeley.edu/~pattsrn/talks/BadTalk.pdf>, 1983.

b D. Patterson, <http://www.cs.berkeley.edu/~pattsrn/talks/BadCareer.pdf>, 1997.

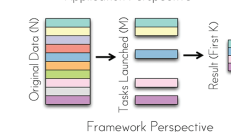
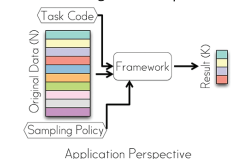
KMN-Computations

Shivaram Venkataraman, Aurojit Panda, Michael J. Franklin, Ion Stoica

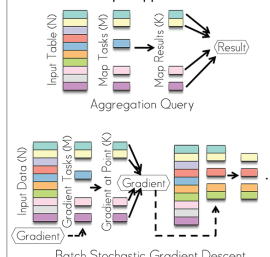
Data doubles every 15 months
Data access speeds limit time to process
- A fundamental limitation
Applications increasingly use samples.



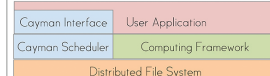
Introducing KMN-Computations



Example Applications



Cayman
Our prototype implementation on top of Spark
Adds a new kind of sampled RDD
Minor scheduler changes.

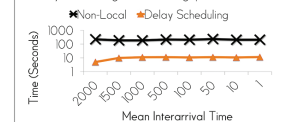


Challenges

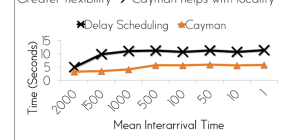
- Framework should determine "M"
- Mitigating Correlation
Fastest K out of M might be data dependent
In-memory blocks more likely to be picked
- Using developer provided sampling function
Currently only supports uniform random.

—amplab

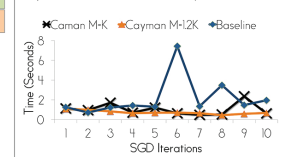
Locality based Scheduling
The rise of memory based frameworks → Locality Relevant Again
Memory has 50x greater throughput than networks



Greater flexibility → Cayman helps with locality



Straggler Mitigation
Commonly performed by launching extra tasks.
Cayman enables redundant computations



A sample poster from the 2013 Berkeley EECS Annual Research Symposium open house at UC Berkeley's AMPLab.

ments on "How to Build a Bad Research Center." I later suggest how to avoid bad centers.

Bad Commandment 1. Thou shalt not mix disciplines in a center. It is difficult for people from different disciplines to talk to each other, as they do not share a common culture or vocabulary. Thus, multiple disciplines waste time, and therefore precious research funding. Instead, remain pure.

Bad Commandment 2. Thou shalt expand centers. Expanse is measured

geographically, not intellectually. For example, in the U.S. the ideal is having investigators from 50 institutions in all 50 states, as this would make a funding agency look good to the U.S. Senate.

Bad Commandment 3. Thou shalt not limit the duration of a center. To demonstrate your faith in the mission of the center, you should be willing to promise to work on it for decades. (Or at least until the funding runs out.)

Bad Commandment 4. Thou shalt not build a graven prototype. Integrat-

ing results in a centerwide prototype takes time away from researchers' own, more important, private research.

Bad Commandment 5. Thou shalt not disturb thy neighbors. Good walls make good researchers; isolation reduces the chances of being distracted from your work.

Bad Commandment 6. Thou shalt not talk to strangers. Do not waste time convening meetings to present research to outsiders; following the eighth commandment, reviews of your many papers supply sufficient feedback.

Bad Commandment 7. Thou shalt make decisions as a consensus of equals. The U.S. Congress is a sterling example of making progress via consensus.

Bad Commandment 8. Thou shalt honor thy paper publishers. Researchers of research measure productivity by the number of papers and the citations to them. Thus, to ensure center success, you must write, write, write and cite, cite, cite. If the conference acceptance rate is 1/X, then obviously you should submit at least X papers, for otherwise chances are that your center will not have a paper at every conference, which is a catastrophe.

Alternatives to a Bad Research Center

Creating alternatives to bad centers requires breaking all eight command-

ments. While I use my Berkeley experience for concrete examples, I polled the alumni from the accompanying table and found that projects at CMU, Google, Harvard, Wisconsin, and UC San Diego are breaking these commandments as well.

Good Commandment 1. Thou shalt mix disciplines in a center.

Good Commandment 2. Thou shalt limit the expanse of a center. The rapid change in underlying technologies, and hence our fields, leads to new opportunities. While industry has more resources, it is often difficult for companies to innovate across conventional interfaces. The psychological support of others also increases the collective courage of a group. Multidisciplinary teams, which increasingly involve disciplines outside computer science, have greater opportunity if they are willing to take chances that individuals and companies will not. I believe in our fast-moving fields there are now more chances for impact between disciplines than within them.

Proposers often promise nearly anything to increase the chances of getting funding, with little regard to running a center should funding be procured. For example, many believe that including numerous faculty and institutions increases chances of funding. One excuse is simply the

difficulty of evaluating research; as it takes time to judge center impact, there is little downside to proposing unwieldy centers. One study suggests following both of these good commandments. After examining 62 NSF-funded centers in computer science, the researchers found that multiple disciplines increase chances of research success, while research done in multiple institutions—especially when covering a large expanse—decreases them: “The multi-university projects we studied were less successful, on average, than projects located at a single university. ... Projects with many disciplines involved excelled when they were carried out within one university.”²

A downside to multidisciplinary research is that it does take time to understand the differences in culture and vocabulary. If you believe multidisciplinary centers offer the best chance for having impact, however, then the benefits outweigh the costs.

Good Commandment 3. Thou shalt limit the duration of a center. My career has been based on five-year research centers, as the table attests. This sunset clause arose from three observations:

- *To hit home runs, it is wise to have many at bats.* Fortunately, people remember research home runs and not the near misses. My experience has been that the chance for home runs is more a function of the number of research projects than of the years spent, so shorter projects give you more chances for success.

- *It is difficult to predict information technology trends much longer than five years.* We start a center based on our best guess of what new opportunities will present themselves in seven to 10 years, which is the right target for a five-year research center. It is much, much more difficult to guess what the opportunities will be in 15 to 20 years, which you would need for longer projects.

- *U.S. graduate student lifetimes are about five years.* It is much easier to run a center if there is no turnover of the people. As we are in academia, at the start of each new project we recruit a new crop of students, as they will not graduate for five years.

You need a decade after a center finishes to judge if it was a home run.

David Patterson's research centers (the research center director is listed first in the third column). The growth in project size over time probably has as much to do with our increasing success at fund-raising as it does with trying to tackle bigger research problems.

Years	Title	Profs: Director, Co-PIs	Students
1977–1981	X-Tree: Tree Multiprocessor	Despain, Patterson, Sequin	12
1980–1984	RISC: Reduced Instructions	Patterson, Ousterhout, Sequin	17
1983–1986	SOAR: Smalltalk On A RISC	Patterson, Ousterhout	22
1985–1989	SPUR: Symbolic Processing Using RISCs	Patterson, Fateman, Hilfinger, Hodges, Katz, Ousterhout	21
1988–1992	RAID: Redundant Array of Inexpensive Disks	Katz, Ousterhout, Patterson, Stonebraker	16
1993–1998	NOW: Network of Workstations	Culler, Anderson, Brewer, Patterson	25
1997–2002	IRAM: Intelligent RAM	Patterson, Kubiawicz, Wawrzyniec, Yelick	12
2001–2005	ROC: Recovery Oriented Computing Systems	Patterson, Fox	11
2005–2011	RAD Lab: Reliable Adaptive Distributed Computing Lab	Patterson, Fox, Jordan, Joseph, Katz, Shenker, Stoica	30
2007–2013	Par Lab: Parallel Computing Lab	Patterson, Asanovic, Demmel, Fox, Keutzer, Kubiawicz, Sen, Yelick	40
2011–2017	AMP Lab: Algorithms, Machines, and People	Franklin, Jordan, Joseph, Katz, Patterson, Recht, Shenker, Stoica	40
2013–2018	ASPIRE Lab	Asanovic, Alon, Bachrach, Demmel, Fox, Keutzer, Nikolic, Patterson, Sen, Wawrzyniec	40

Just eight of the 12 centers listed in the table are old enough, and only three of them—RISC, RAID, and the Network of Workstations^c center—could be considered home runs. If slugging .375 is good, then I am glad that I had many five-year centers rather than fewer long ones.

A downside to sunset clauses is that it is easier to recruit students and sustain funding as a center grows in reputation than to recruit to or to fund a new center. However, if the goal is to maximize home runs versus recruiting or funding, in our fast-moving fields it is better to declare victory after five years and look afresh for new opportunities, and then to recruit the right team to pursue them.

Good Commandment 4. Thou shalt build a centerwide prototype. A common feature of the centers in the table is the collective creation of a prototype that demonstrates the vision of the center, which helps ensure the pieces are compatible. While systems students like building their part of the center vision, they do not necessarily want to work with others to make everything fit together. However, the educational power of such multidisciplinary centers comes from students of different backgrounds working together, as the experience improves their understanding and taste in research. This process also enhances students' system-building skills by requiring them to do a real prototype, not just toy demos. Such prototypes can even lead to open source projects, which simultaneously aid technology transfer and expand the workforce building the prototype. In fact, open source success generally means more developers outside versus inside the center.

One downside of such centers is that faculty must convince students of the benefits of spending some of their time working for the common good versus working just on one's own piece. However, once they start seeing how the research of others benefits their results, they no longer require encouragement to do so.

^c The NOW project showed clusters of workstations helped everything from encryption to sorting. The Inktomi search engine was built on NOW. The startup Inktomi Inc. in turn proved the value of the value of clusters of many low-cost computers versus fewer high-end servers, which Google and others in the Internet industry later followed.

I believe in our fast-moving fields there are now more chances for impact between disciplines than within them.

Good Commandment 5. Thou shalt disturb thy neighbors. Researchers have found that innovation is enhanced if all participants work in a single open space less than 50 meters across, as it encourages spontaneous discussions across disciplines.¹ The goal is for the space to support concentration and communication. For example, for the last four centers listed in the table, the faculty gave up their private offices to be embedded with students and postdocs in open space, where only the meeting rooms have walls.⁸ Faculty access draws students from their home offices to the lab, which increases chances of interactions.

A downside of shared space is the cost of remodeling to create an attractive open space. This is a one-time capital expense, since following projects will use it, but even so, the cost was equal to just two students over the life of a center. Shared open space is certainly more beneficial to a center than a few more students.

Good Commandment 6. Thou shalt talk to strangers. A key to the success of our centers has been feedback from outsiders. Twice a year we hold three-day retreats with everyone in the center plus dozens of guests from other institutions. You can think of it as having a "program committee" that meets with everyone for six days a year for five years, or a month of feedback per center. Their value is indicated by 100% of the respondents in my alumni poll using them.

Having long discussions with outsiders often reshapes the views of the students and the faculty, as our guests bring fresh perspectives. More impor-

tantly, at the end of each retreat we get frank, informed, and thorough feedback from our valued guests. Researchers desperately need insightful and constructive criticism, but rarely get it. During retreats, we are not allowed to argue with feedback when it is given; instead, we go over it carefully on our own afterward. In fact, we warn that we are taking their advice seriously, so be careful what you ask for.

Retreats fulfill other important roles:

- They provide real milestones, which are scarce in academia. It is one thing to promise your advisor you will get something done by January, but quite another when you are scheduled to give a talk on it to respected visitors.

- They build esprit de corps, in that everyone in the center spends three days together. There is free time to play as well as work, so students in these centers often become lifelong friends. In fact, I was delighted to learn from the poll that perhaps due to their shared experience, there is even a sense of connection between the generations of systems alumni.

- Over their five-year center lifetimes, all students get 10 chances to present posters or give talks, which gives them as much exposure as they desire and improves their communication skills.

A downside to retreats is again cost, which is about as much as one or two students. As before, exchanging retreats for a few more students would be ill advised. You also need to be careful which outsiders you invite to the retreat to be sure that they are really engaged and that they will offer useful and constructive criticism.

Good Commandment 7. Thou shalt find a leader. To make progress, you need to find someone willing to spend the time to pull people together, to create a shared vision, to build team spirit, and in who investigators believe will make decisions in the best interest of the center. When I am the director, I am playing the game to be on the winning team rather than to be the most celebrated coach. Hence, I recruit faculty who are team players, leaving prima donnas for others to manage. I try to lead by example, working hard for the center's success while hoping my teammates are inspired to follow. My guiding motto is "There are no losers on a winning team, and no winners on a losing team."

Good Commandment 8. Thou shalt honor impact. While papers are surely associated with high-impact research, they are not a direct measure of it. High-impact research can even be difficult to publish. The first paper on the LLVM compiler, winner of the 2012 ACM Software Systems Award, was initially rejected by the main compiler conference (PLDI), and the first paper on the World Wide Web by Tim Berners-Lee was rejected by Hypertext '91.

Nor are best paper awards necessarily predictors of high impact. One study counts citations of best paper awards, documenting the modest influence of most of them.^d For example, the conference program committee for the first MapReduce paper, which led to the 2012 ACM Infosys Award, named two other papers as best! Test of time awards, given 10+ years after the conference, are a much better indicator of success but they are trailing indicators.

For my successful projects, the early indicators were not papers or awards, but non-researchers trying to use our ideas. For RISC and NOW, the first adopters were startups or at least young companies. For RAID, it was hungry companies trying to gain market share. The market leaders were happy with the status quo; they developed related products only *after* others had commercial success.

Educational Impact of Centers

Some may be surprised to learn that centers can improve undergraduate education, which is a principle of research universities. Less than a decade after seeing how to pipeline microprocessors in the RISC project, universities taught undergraduates to design their own. The RAD Lab led to a reinvention of our software engineering course, which has long been problematic in CS departments.³ It grew quickly from 35 to 240 Berkeley students, and inspired massive open online courses where 10,000 earned certificates in 2012. These classroom innovations led to three textbooks, which further expand educational impact.^{4,5,9}

Good research centers are really

federations of subprojects that share a common vision, with each subproject having one or two professors and several students. Each professor is an expert in a different field, which facilitates multidisciplinary research. These subprojects give students the same faculty attention as single-investigator project, but in addition students:

- Learn by working and negotiating with dozens of smart, hard working, and stubborn collaborators, some of who go on to notoriety.

- Acquire good taste in research that comes from working on a common prototype, which leads to clear-eyed dissertations.

- Get a user base and feedback for new tools and approaches.

- Have senior students as role models in the open space, which can lead to explicit mentoring.

- Have fewer feelings of loneliness that is all too common in the Ph.D. process, due to the esprit de corps that comes from the open space and retreats.

- Can be reenergized by praise from retreat visitors.


I cannot be modest while espousing the successes of this model, so as I warned earlier, let me yet again apologize in advance. When DARPA cut the research funding to universities in the last decade,⁷ our pitch to companies was that we needed multidisciplinary centers to keep producing sterling systems students. Happily, our claim proved to be true. For example, one student from the first cohort in the open space of the RAD Lab came up with the idea for the Spark cluster-computing framework¹¹ after overhearing machine-learning students in the lab gripe about MapReduce. Not only did he get job offers from all companies he visited, he got them from all the top universities too.

In case you were curious, these projects do not just produce single superstars. For example, the Par Lab averaged three students per paper and each student averaged three first-papers. More importantly, Burton Smith opined: "These are the best graduate students I have ever seen." This is high praise from a Microsoft Fellow and National Academy of Engineering (NAE) member. What Smith said in 2013 was obviously gratifying, but I was struck

that this is exactly what Mark Weiser, the father of ubiquitous computing, said about SPUR students in 1989.

Reflections on Five-Year, Multidisciplinary Research Centers

If such centers are good for faculty as well as for their students, then by now we should be able to tell. One measure is election to NAE and the National Academy of Sciences: only five to 10 computer scientists are selected each year for NAE (with half from industry), and just two or three for NAS. Thus far, nine of the 24 Berkeley CS faculty who worked in the centers listed in the table are in NAE (~40%), and three are also in NAS. In fact, the majority of the Berkeley CS faculty in NAE or in NAS worked on these projects. In case you were wondering, they were elected after joining these centers.

Whereas early computing problems were more likely to be solved by a single investigator within a single discipline, I believe the fraction of computing problems requiring multidisciplinary teams will increase. If so, then learning how to build good centers could become even more important for next 40 years than it has been for the past 40. 

References

1. Allen, T. and Henn, G. *The Organization and Architecture of Innovation: Managing the Flow of Technology*. Butterworth-Heinemann, 2006.
2. Cummings, J. and Kiesler, S. Collaborative research across disciplinary and organizational boundaries. *Social Studies of Science* 35, 5 (2005), 703–722.
3. Fox, A. and Patterson, D. Crossing the software education chasm. *Commun. ACM* 55, 5 (May 2012), 44–49.
4. Fox, A. and Patterson, D. *Engineering Software as a Service: An Agile Approach Using Cloud Computing*. First Edition. Strawberry Canyon, 2014.
5. Hennessy, J. and Patterson, D. *Computer Architecture: A Quantitative Approach*. Fifth Edition. Morgan Kaufmann, 2011.
6. *Innovation in Information Technology*. National Research Council Press, 2003.
7. Lazowska, E. and Patterson, D. An endless frontier postponed. *Science* 308, 5723 (2005), 757.
8. Patterson, D. Your students are your legacy. *Commun. ACM* 52, 3 (Mar. 2009), 30–33.
9. Patterson, D. and Hennessy, J. *Computer Organization and Design: The Hardware/Software Interface*. Fifth Edition. Morgan Kaufmann, 2013.
10. *The Berkeley Par Lab: Progress in the Parallel Computing Landscape*. D. Patterson, D. Gannon, and M. Wrinn, Eds., 2013.
11. Zaharia, M. et al. Spark: Cluster computing with working sets. In *Proceedings of the 2nd USENIX Conference on Hot Topics in Cloud Computing* (2010).

David Patterson (pattsrn@cs.berkeley.edu) is the E.H. and M.E. Pardee Chair of Computer Science at UC Berkeley and is a past president of ACM.

Copyright held by Author/Owner(s).

An interview with David Patterson appears on page 112.

^d See Best Papers vs. Top Cited Papers in Computer Science; <http://arnetminer.org/conferencebestpapers>.