Season 3, Episode 2

Margaret Lawson: Finding Her Place

SPEAKERS
Sarah Webb, Margaret Lawson

Sarah Webb 00:00
I'm your host, Sarah Webb, and this is Science in Parallel, a podcast about people and projects in computational science. In this episode, I'm speaking with Margaret Lawson. She's a software engineer working with Google on cloud storage from the company's Kirkland, Washington, location. Margaret's high-performance computing career launched at Sandia National Laboratories through an undergraduate internship and later through her doctoral research, where she tackled challenges in managing metadata. Margaret completed her Ph.D. at the University of Illinois, Urbana-Champaign, supported by a Department of Energy Computational Science Graduate fellowship.

Sarah Webb 00:45
Initially, Margaret didn't expect to build a career in computing. We talked about the mentors who encouraged her, her pursuit of challenging problems, progress for women in computing and her work to facilitate discussions about ethics and HPC. We recorded this conversation last November at the supercomputing meeting, SC22, in Dallas, Texas.

Sarah Webb 01:08
Margaret, it's great to have you on Science in Parallel.

Margaret Lawson 01:10
Thank you for having me, really appreciate it.

Sarah Webb 01:12
I'd like to start by learning a little bit about how you got your start.

Margaret Lawson 01:18
I've always been interested in math and science. I come from Davis, California, a pretty small university town, and my sister, my parents, they're all biologists. So science has always been really big in my family. But in college, I actually came in as an econ major, and pretty quickly realized maybe econ wasn't for me. And then I thought maybe psychology would be my major. And I ended up actually taking my first computer science course to meet a distributed requirement. And because I thought it might help with trying to get into like psychology research labs, and I actually ended up loving it. And the rest, as they say, is kind of history.

Sarah Webb 01:56
Wow. So at what point did you decide that you wanted to pursue graduate work?

Margaret Lawson 02:05
That's another one where I kind of just lucked out and walked into it. So my undergraduate school, Dartmouth, is a little weird. They make you take classes your sophomore summer, and then you take off some other quarter. So I ended up taking off my junior spring and summer, so I had like a five month chunk off. And I really wanted to do an internship in that time, and ideally, an extended internship. And I started looking around, and I knew someone who had worked at Sandia National Labs for decades and absolutely loved it. And I said, Do you know if anyone's looking for computer science intern? And someone was, and someone was specifically trying to get more women into the national lab that was looking to have a female intern.

Sarah Webb 02:46
That someone was Jay Lofstead.

Margaret Lawson 02:49
He said," This is going to be incredibly difficult and hard work and confusing. I'm gonna work your way too hard. How does that sound?" And I said, "Great." And I ended up really loving my time there and really loving what I was working on and grad school was the way I could keep doing it.

Sarah Webb 03:02
So what did you work on during that time?

Margaret Lawson 03:05
I was working on metadata management for high-performance computing. So basically, looking at all these HPC scientists have way too much data, and they have trouble sometimes finding what they think is interesting, or keeping track of that, and what can we do to help?

Sarah Webb 03:20
So how did that then lead to graduate work for you?

Margaret Lawson 03:24
I mean, it's a pretty open-ended research question. It's graduate work in that it's research. And that's really what PhD was all about is teaching you how to be a good independent researcher. So I had already developed some tools for how to do this, and grad school was the way for me to get better at it.

Sarah Webb 03:41
So what were you working on in your Ph.D.? Was it an extension of those sorts of questions?

Margaret Lawson 03:48
It was very much an extension. Really, what I did as an undergrad kind of ended up being the groundwork for what I did in my Ph.D. And it was the initial exploration that taught me these are some good ideas that need to be changed or built upon or modified in different ways. So it really helped guide

me in getting deeper with the work and also creating what I think is a better system ultimately because I was a systems researcher.

Sarah Webb 04:10
So walk me through a bit of your PhD. How were you redesigning the system to make it better?

Margaret Lawson 04:17
HPC is kind of a really interesting field, right? Because compute power of supercomputers has really exploded in recent years, and the amount of data produced by them has kind of exploded as a result. So it's really only pretty recently that I think metadata has become so important. Because once your metadata starts getting on like the gigabyte, terabyte, maybe petabyte scale, all of a sudden, you need to start looking at a lot of these more traditional, like data management challenges that have been looked at for years. But now it's our metadata is on that scale.

Margaret Lawson 04:49
 So a lot of what I was looking at was, how can we manage this large amount of information efficiently? No, again, HPC performance, scale is everything. So those were some, you know, just basic principles for any system developed had to be for this environment had to be able to meet these requirements. And then just often the scientists in HPC are not computer science experts, they really want something that's usable, intuitive that really does what they want without them having to be experts in the system or develop their own system from scratch every time. So the very initial design that we had was a system that had basically you spin up nodes that were dedicated for metadata management. So anytime you wanted to write or query the metadata, you would send a message to these nodes. And while that can be really good for like, availability purposes, it's not the best from a performance perspective because metadata in the realm of is really high bandwidth networks is small messages, sending a large number of small messages across the network really was not as performant as we wanted it to be.

Margaret Lawson 06:02
So ended up I developed a local version where you could have basically your metadata servers embedded within the compute nodes as an option and then ultimately ended up having that be the design of the system. So that was one of the ways. And then another kind of limitation of the initial system was it started by tackling the easiest case, which is applications that use regular meshes. So it's pretty simple if you're using a regular mesh, and a scientist says, Hey, from these latitude and longitude coordinates, there was something interesting here is really easy, then go back and say, Okay, here's the exact data that's involved in this feature,. It gets a lot more complicated when you're using an unstructured mesh, because you can't make any assumptions about the mapping of like, what data is actually in that space in your application in your simulation. So the system was working for my Ph.D. was expanding it to applications that had more complex mesh types, including completely unstructured meshes.

Sarah Webb 07:02
Did you continue to work with Sandia throughout your Ph.D. on that? Or how did that work?

Margaret Lawson 07:07
Yeah, I did. And I'm really grateful that I was able to do that. So in my undergrad, I started working at Sandia and had a fantastic mentor, Jay Lofstead, who basically continued to mentor me throughout my entire Ph.D. So I was on-site, junior year, spring, summer for five months as a full time intern. And then from there on till the end of my Ph.D., I was a year-round intern at Sandia. And so what that did was, you know, enabled me to continue being mentored, which was fantastic. It also gave me continuing access to the supercomputers, which were really vital for me to be able to develop a system like when you're designing something, again, for HPC workloads, you need a lot of cores to develop the system, you need a lot of cores to test the system. And that can be something that it's really hard to get access to as a student. That was really invaluable. And it also just meant that I had access to be able to talk to all of these scientists to say, What challenges are you having? What do you need? How can I help?

Sarah Webb 08:02
So not only the resources to solve the problem, but the people who really could tell you, this is the problem that I need you to solve?

Margaret Lawson 08:09
I mean, the joke, and it's absolutely true in systems research, is you often have people developing systems that no one wants, no one needs, no one's going to use. The question is, how can you try and make sure that you're actually solving the problem you set out to solve? And I think one of the most important ways of doing that is talking to the people you're trying to solve it for. I think often, the computer scientists and the application scientists do kind of speak a different language. And so a lot of my learning process as an undergrad as a graduate student is learning enough with a language that I could communicate effectively.

Sarah Webb 08:48
What did you do for your practicum, as part of the CSGF?

Margaret Lawson 08:51
I ended up going and working with Lawrence Livermore, a sent an *in situ* analysis and visualization library for HPC. So I'd mostly in the past been developing these very like beta metadata management-heavy systems. And this was a chance to look at another piece of the general workflow, which is once you're done actually running the simulation, once you're done storing the data analysis and visualization is usually what you're doing with the data like what comes next, or in the case of in situ, what comes from the middle?

Margaret Lawson 08:53
What were your takeaways from doing that? Because I'm sure it drew on a lot of core skills that you already developed, but you learned some other ones, I'm sure.

Margaret Lawson 09:34
I think part of the advantage is I had largely been working with the same group of people for last three, four, five years. And it was a good chance for me to get to work with a different group of people. And I think you learn a lot on every team that you're in. And I think also to the way I saw my job, as a

researcher was to try and better understand the entire lifecycle of an HPC application, the entire workflow of HPC scientists and this was just insight into another piece of that.

Sarah Webb 10:06
I'm going to shift gears for a second because I want to ask you a little bit about a you were organizing this Birds of a Feather (BoF) session on ethics and HPC. Can you talk a little bit about that session, and maybe some of the backstory of how that session got started, and some of the key issues that came out of the discussion?

Margaret Lawson 10:28
Yeah, one of the things that I absolutely loved about my undergraduate degree was I went to Dartmouth, a pretty small liberal arts college, and they kind of lovingly force you to try different things. And yeah, they have these distributed requirements. And again, that's how I got into computer science. And then also as a different distributed requirement. I took a political philosophy course. And it turns out, I love that,too. And I ended up actually taking a lot more ethics courses while I was there. And, you know, from Dartmouth, I had, like, many professors I loved. One of my absolute favorites was Tom Cormen, a computer science professor, and the other one was Julie Rose, an ethics professor. So it was really two of the absolute favorite things I studied in undergrad. And I was sitting at SC in 2018, with Jay Lofstead, and Jakob [Lüttgau], so the other two people I'm on the BoF with, and also Elaine from Sandia.

Sarah Webb 11:27
That's Elaine Raybourn. She was a guest on season two of this podcast,

Margaret Lawson 11:32
We were kind of talking and as it sometimes does, when you're talking with friends, we start talking about some of your favorite things. For me, that was ethics. And we were getting into some debates, and were saying it would be good to be able to continue these discussions. And none of us had really seen much HPC-focused conversations about ethics. And we really just wanted the opportunity to have that community discussion if we could or start the conversation. And so we ended up doing a BoF at SC19, the next year and have been able to continue it three out of the last four years.

Sarah Webb 12:03
What would you say have been the takeaways, so far, from that discussion? What do you think the community is thinking about? And what else do you think the community should be thinking about in terms of ethics?

Margaret Lawson 12:16
I think honestly, the big problem we run into is this is one hour out of the year, and then you wait another year, and then you get one more hour. And it's great to have community support. It's great to have people engage in discussion, but I think we need more. And a lot of what we're looking at now is how do we do that? But I think lots of really good discussions ranging from everything from what do the creators of a technology, what responsibility do they have to try and manage it and manage how it's used? The BoF today talked a lot about like, global equality in terms of access to these resources, but

then, you know, it gets all the way down into not just what countries have access, but what maybe institutions or companies within a country have access. And then I think we talked a little bit in the BoF, too, about what individuals within a society have access and who gets to show up at these conferences and who gets to show up in this community. So and then that's not even touching on like the environmental considerations, particular applications like national defense, or oil and gas, or all of these things. There are just so many topics for ethical considerations, I think, for any community, but in particular, for this compute community because of the scale at which we do things, and the industries in which the technology is being used.

Sarah Webb 13:39
So I know at the end of that session, there was kind of a wrap up and some discussion about where things might go next with this discussion of HPC and ethics in the future. Do you think you can get this expanded to a workshop or an ongoing discussion? What are your goals there anyway?

Margaret Lawson 13:56
Yeah, I mean, I think, for me, something like a workshop would be part of the goal. I think, one challenge we have is, you know, in general, people who are here at this conference may not be ethics experts. They're here because they're HPC experts. So I think having the opportunity to bring in outside experts would be really useful. And I think more like having more time with a workshop would maybe allow us to get into providing more information, providing more background, yeah, continuing with the discussion component. And then we're talking a little bit about the BoF today, and maybe having people write position papers, so you're not having that time constraint so that people can really dig more into the details in the meat of issues.

Sarah Webb 14:42
It sounds like this is a discussion that you're going to plan to be continue to be involved in. I assume that's your goal. Is this kind of a side passion project for you?

Margaret Lawson 14:50
Yeah, absolutely. It's always hard to find enough time for everything we're passionate about and everything we're interested in. But yeah, I really enjoy the discussions, and I think they're important to have.

Sarah Webb 14:59
You will alluded to this when you were talking about your practicum. But I do feel like as a society, and then as individual scientists, people, the last two years have been challenging. What do you think are the key lessons that you learned about the last two years that, you know, are shaping you as a scientist?

Margaret Lawson 15:17
I think one of the main takeaways from me and a lot of scientists was just the power of collaboration and having that infrastructure for sharing science, sharing what we've learned in place is just incredibly valuable. Like, that's how we came up with vaccines, right? I think COVID, more than maybe any other major health crisis has shown scientists coming together and working together, like at their best. And I

felt incredibly grateful that the research that I do can be done remotely, because I did have many friends, who were fortunate enough to be employed as graduate student researchers, but they weren't able to get into their labs for often many months. You know, some of them were in biology, and you have things like cell cultures, and you can't be there to take care of them, you kind of have to throw them away and start over. So there were many researchers who, for a long time, weren't able to do their work. So I felt really grateful that I could.

Sarah Webb 16:19
Tell me about your new job at Google. What are you working on?

Margaret Lawson 16:22
I am working on expanding Google Cloud storage offerings. DAOS is an object storage system, it's often doing very well on like I/O 500 lists, for example, being used by Aurora. And so for this conference, I was able to put together like a DAOS demo, just showing how well it performs on Google Cloud's platform, and really useful for a lot of HPC workloads that involve like a large number of files, or a large number of small files. And in particular, the demo I was looking at was so a deep learning use-case, which again, I think you often have those large number of very small files. So just very useful to be able to access those more quickly and just help people accelerate their science.

Sarah Webb 17:16
Is there anything that we haven't talked about that you think is important to mention?

Margaret Lawson 17:21
I feel very grateful, very lucky to have ended up where I did. So I mean, again, kind of stumbled into computer science by accident and then even once I did, you know, I really loved it, but wasn't sure. If I had a place in the field, I think. For all kinds of reasons that we don't have time to get into I had this, like, long standing impression from being a kid that's like, oh, no, computer science wasn't for me, like, no computers weren't for me. And I actually had a male mentor saying, "Oh, no, of course, you're going to major in computer science. Like, of course, you're going do it. Like, of course, you should consider going to grad school." And so he really helped me feel like I had a place in the field.

Margaret Lawson 17:59
And then again, I, you know, wasn't sure about grad school, I wasn't sure what I was doing next. And then I, you know, had Jay Lofstead saying, "No, we need more women at the labs. We need more women in computer science. We need more women doing this kind of research." and again, kind of helped me find a place there and and, you know, helped me apply to grad school and mentored me ever, ever since then. So again, I just felt really lucky that I had these opportunities, had these mentors and you know, had, yeah, male mentors helping to get more women into this field.

Sarah Webb 18:34
So is it getting easier for women? I guess, where are we and where can we continue to improve?

Margaret Lawson 18:39

I think, you know, I kind of view it as a trickle-down effect. So when I was at Dartmouth, we had actually achieved 50/50 parity in engineering. And we were, I think, around that in computer science as well. But then still, once you get into graduate programs, like particularly Ph.D. programs, I think the best ratio I saw was like 33% female, and a lot of them are closer to 20% or lower, it was very kind of shocking to me in my PhD program. I was the only American woman in my year, I think it was 65 or 67 people. So I think trickle -own effect into grad school. It's, we're still working even close to parity. And then again, if you look one step further and look at professors, that's going to be even later. So it's just made great progress that, you know, I, I, in none of my computer science classes in undergrad was I ever a minority, really. Getting it into the national labs, getting it into the PhD level, getting it going everywhere else, it's still a work in progress, but it's coming along.

Sarah Webb 19:45
That's good. That's good to hear. And is there a single piece of advice that you would pass along to an early career researcher who's thinking about getting started in computer science?

Margaret Lawson 19:59
I would say, so if you haven't tried it, you should try it. Again, I had all of these preconceived notions about what computer science actually looked like, and they were all wrong. And it wasn't until I actually did it and learned what it was and, you know, started coding that I found out that it was something I loved. And then if you're just getting started, and you're not sure if the field has a place for you, I would say it absolutely does. And, again, you should never let your fears about maybe not fitting in or being the only one or not being good enough at it. Or, you know, the fact that some people started coding when they were five years old. You should never let those things stop you. And if you love it, you should do it. If you're not sure, you should try it.

Sarah Webb 20:48
Margaret, thank you so much for your time today. It was such a pleasure talking with you and learning about your work.

Margaret Lawson 20:54
Thank you again, I really appreciate the opportunity. It's been delightful chatting with you.

Sarah Webb 20:59
To learn more about Margaret, her research and ethics in HPC. Check out our show notes at scienceinparallel.org.

Sarah Webb 21:05
Science in Parallel is produced by the Krell Institute and is a media project of the Department of Energy Computational Science Graduate Fellowship program. Any opinions expressed are those of the speaker and not those of their employers, the Krell Institute or the U.S. Department of Energy. Our music is by Steve O'Reilly. This episode was produced by Sarah Webb and edited by Tess Hanson.