HPC User Forum Update

Interviews with Steering Committee Members: David Martin, Argonne National Laboratory

Steve Conway and Thomas Gerard
August 2020

IN THIS UPDATE

After the global pandemic forced Hyperion Research to cancel the April 2020 HPC User Forum planned for Princeton, New Jersey, we decided to reach out to the HPC community in another way – by publishing a series of interviews with members of the HPC User Forum Steering Committee. Our hope is that these seasoned leaders' perspectives on HPC's past, present and future will be interesting and beneficial to others. To conduct the interviews, Hyperion Research engaged insideHPC Media. We welcome comments and questions addressed to Steve Conway, sconway@hyperionres.com or Earl Joseph, ejoseph@hyperionres.com.

This interview is with David Martin, Manager, Industry Partnerships and Outreach at the Argonne Leadership Computing Facility at Argonne National Laboratory. He works with industrial users to harness high performance computing and take advantage of the transformational capabilities of modeling and simulation. Mr. Martin brings broad industry and research experience to ALCF. Prior to joining ALCF, Mr. Martin led IBM's integration of Internet standards, grid and cloud computing into offerings from IBM's Systems and Technology Group. Before IBM, Mr. Martin managed networks and built network services for the worldwide high-energy physics community at Fermilab. He began his career at AT&T Bell Laboratories, doing paradigm-changing work in software engineering and high-speed networking. Mr. Martin has a BS from Purdue and an MS from the University of Illinois at Urbana-Champaign, both in Computer Science.

The HPC User Forum was established in 1999 to promote the health of the global HPC industry and address issues of common concern to users. More than 75 HPC User Forum meetings have been held in the Americas, Europe and the Asia-Pacific region since the organization's founding in 2000.

DAVID MARTIN INTERVIEWED BY DAN OLDS, insideHPC

Olds: This is Dan Olds on behalf of Hyperion Research and insideHPC and we are interviewing David Martin today from Argonne National Lab. How are you doing today, David?

Martin: I'm fine. How are you?

Olds: I'm okay. And this is going to be interesting. I'm looking forward to hearing what you have to say. So, let's start with some basics. How did you get involved in HPC?

Martin: My background really is in computer networking. You know, the dirty secret about HPC, especially high-end supercomputers, is that they're really networks. The processors are a lot of times commodity processors, but the secret sauce is the network.

Olds: Some say it's a dark art.

Martin: It is a bit. But I actually started at Bell Labs after college and did packet communications at a time when packet communications in Bell Labs was a strange thing. They believed everything should be circuit-switched. So, we were a rag-tag group pushing packet-switching and it got me really hooked on the bug of making networks as fast and as interoperable and performant as possible.

Olds: So, would you summarize your HPC career by saying you are a network-based practitioner?

Martin: I guess so. After Bell Labs I moved to Fermilab and I ran wide area networks for quite a while. Then I went to IBM and they were just getting into the Internet, which seemed strange at the time, but at the time they didn't have much of an Internet business or presence at all. They were looking for people who knew TCP/IP and knew how to build networks and so I did that. And, then, networking led me into computing, which is kind of how I got into grid computing. So, I did some grid computing when I was at Fermilab. But a lot of it was going on then at IBM because they really saw networking as the way that they could tie together all of their big computing centers, which up until that time had been pretty much islands. Even within computing standards, all the machines were islands and they wanted to start becoming more of an Internet company and hooking those together. So, that kind of led me to HPC in a roundabout way.

Olds: Okay. What are some of the biggest changes that you've seen in HPC over your career?

Martin: Going back to the days when I was at Bell Labs, I had one of the first Sun Microsystems workstations on my desk. Mostly the reason I had it is because nobody else really could figure it out because they were still using a lot of minicomputer or mainframe-based computing.

Olds: A lot of terminals and things like that?

Martin: Exactly. So, the group that I was part of at Bell Labs there started using them and when I got to Fermilab the boss that I had there was into distributed computing. I think the motto for Sun for a long time was, "the network is the computer."

Olds: Yes. I was actually with Sun during those glory days.

Martin: Oh, interesting. It was fun. It was a small enough company that when I found bugs, I would call the engineer that wrote the code and let that person know. You don't get to do that much anymore.

Olds: No, no you don't. Very interesting stuff.

Martin: Just to finish the question about some of the changes I've seen, I think it's taken a long time for those changes to propagate through the whole computing sphere. So, that was big among people who were using Sun Microsystem stuff, but it took a while. If you look at a lot of the cloud computing now, again, the network is the computer. The ability of cloud providers to provide all this raw computing power is really dependent upon having networks that work and are reliable and interoperable and all that. So, I think distribution through the whole computing ecosystem is the big change. That, and the fact that now people are really having to do parallel programming. You could get away, for a really long time, with single-threaded code and ignoring the fact that there are multiple nodes and all that. You can't get away with that now. Even if you're programming for your laptop PC you have 16 cores or something like that in it now.

Olds: And if you program for single-thread, your performance is just going to go down over time.

Martin: Right. Even in the supercomputers now that the clock-rates are not going up a huge amount. It's all about being able to take advantage of parallelism.

Olds: Yes, exactly. So, where do you see HPC headed in the future? Are there any trends that excite you or have you concerned?

Martin: Well, the introduction of accelerated computing in GPUs, I think, is both exciting and concerning. It's exciting to have that much access to raw computing power. We haven't been able to have a step change like that in quite a while. We've been limping along with trying to even hang onto Moore's Law. There's now the ability to stick a whole bunch of GPUs in a machine. The Aurora system that's coming to Argonne next year, most of the computing power will be in the GPUs. So, having an exascale level of computing power is really exciting. The challenge is what it'll mean for the codes that are out there now. Refactoring all those codes is non-trivial. I'm involved a bit in the exascale computing project and it's a huge effort to refactor a lot of codes that are important to the Department of Energy and the general scientific community and make them work in an exascale environment. But that's a lot of work. It's exciting and it's interesting, but I'm worried somewhat about all of the accumulated HPC code that's around and maybe not working great in the GPU environment.

Olds: That is going to take a lot of time and effort and expense. But, at least, I think, we are fairly certain that GPUs are going to be around, so you're not doing it for nothing.

Martin: I hope so. All that work for nothing would not be good. The other trend that I see that, I think, will probably help is coming from the cloud world, which is virtualization and being able to do things like Docker, Shifter, Kubernetes, and all that. You haven't seen those too much in the HPC environment, but they're starting to go in there and I think the next generation of supercomputers will natively support that and that'll help a lot. Maybe people won't necessarily have to rewrite every application for every piece of hardware the way that they do now in HPC if you have virtual machines. Most of the rest of the computing world has been willing to trade off performance for programmability. So, you know, a lot of programs aren't, maybe, that efficient but they're portable. You can run them on all kinds of hardware. HPC has always been about squeezing the last ounce of computing out of a piece of hardware and getting everything out of the way. I think the computing power is starting to be there where you do not have to worry about that so much.

Olds: There's always been a worry with virtualization about the overhead of it.

Martin: It's true. There is overhead. But even if you look at traditional HPC applications, a lot of them don't run at the raw computing power that is there anyway. Some do, some people pride themselves on the fact that they're getting 70-80% of peak.

Olds: It's going to be interesting. I think, when we get into the low single-digits of overhead that that'll mark the sea-change toward virtualization.

Martin: Yes, and as in the cloud there'll just be so much raw computing power that it'll be worth the trade-off.

Olds: I think that makes sense. Is there anything else you want to say that we didn't cover that I should've asked?

Martin: Well, I think the challenge in high-performance computing, especially as cloud comes along and there are more and more capabilities in commercial offerings, is for high-performance computing to make the systems usable for a wide variety of applications and environments and people. We were joking at the very beginning that the large HPC centers look down on other people and it's kind of a joke, but the reason for that perception is that a lot of these systems are just fiendishly hard to use. So, it requires a cadre of people who have spent years building it and designing it and understanding it. So, I think the challenge for the high-performance computing community in general is to open that up and make it so that a lot of people can make use of it. I think that the emergence of the exascale systems will help that some because the design of them will be much more open. We'll be able to do data analysis, artificial intelligence, traditional simulation, all simultaneously.

Olds: You have a fair amount of capacity then?

Martin: Yes, we can fill it up. We've never run into a situation where we've got a new machine and we don't know what to do with it.

Olds: Yes. That's something that, when I was on the vendor side, we would come out with bigger machines and people say, "Well, it doesn't need to be that big." Yeah, it does. It'll get used.

Martin: We've already got projects that basically are asking for the whole Aurora system. Pieces haven't even started showing up yet. We're a year away from even seeing hardware.

Olds: But they're already pitching you on using the whole thing?

Martin: They're already pitching on why these projects need that amount of time.

Olds: Wow! Well, I look forward to hearing, in the future, what those projects were because that's a lot of capacity to use.

Martin: If you look at a lot of the traditional problems, like astrophysics or climate, today you just shrink your grid a little bit or change your resolution.

Olds: Add some more variables and it starts really ramping up.

Martin: It's easy to get exponential growth in computing power as well. And, you know, some of the really creative things that they're doing will actually require it. It's not just that they're just trying to fill up the computer.

Olds: Well, great. Thank you so much for the time, this has been really good. Really appreciate it and on behalf of Hyperion Research and InsideHPC, thank you all for watching and we'll talk to you again soon.

About Hyperion Research, LLC

Hyperion Research provides data-driven research, analysis and recommendations for technologies, applications, and markets in high performance computing and emerging technology areas to help organizations worldwide make effective decisions and seize growth opportunities. Research includes market sizing and forecasting, share tracking, segmentation, technology and related trend analysis, and both user & vendor analysis for multi-user technical server technology used for HPC and HPDA (high performance data analysis). We provide thought leadership and practical guidance for users, vendors and other members of the HPC community by focusing on key market and technology trends across government, industry, commerce, and academia.

Headquarters

365 Summit Avenue
St. Paul, MN 55102
USA
612.812.5798
www.HyperionResearch.com and www.hpcuserforum.com