

HPC User Forum Update

Interviews with Steering Committee Members: Steve Conway, Hyperion Research

Thomas Gerard and Jean Sorensen
July 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 Steve Conway, senior advisor for HPC market dynamics at Hyperion Research. He produces opinion pieces, studies and reports on the worldwide HPC market, especially in the areas of AI and high performance data analysis, cloud computing, edge computing and the IoT. Steve works closely with government agencies, industry and academia, and the vendor community in North America, Europe and Asia. He was vice president of investor relations and corporate communications for Cray and had management roles at SGI and CompuServe. Earlier, Steve had a 12-year career in university teaching and administration at Boston University and Harvard University. A former Senior Fulbright Fellow, he holds bachelor's and master's degrees in German from Columbia University and a master's in comparative literature from Brandeis University, where he also completed doctoral coursework and exams.

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.

STEVE CONWAY INTERVIEWED BY DAN OLDS, insideHPC

Olds: Dan Olds here on behalf of Hyperion Research and InsideHPC. We're going to have an interview with Steve Conway, who has had a storied career in HPC. So, how did you get involved with HPC in the first place?

Conway: My work in college and grad school was in literature, but I worked with computers a lot, starting with a night job as a system operator for an IBM 360 machine at an insurance company. At a later job, I was a system administrator for a mini-computer multi-user system, so I was always interested in computers.

Olds: And, yet, you were taking literature classes?

Conway: Yes, I majored in literature and minored in physics, so I was interested in science, too. Then, I had a consulting job where Cray Research became one of my very important clients. So, I got involved with HPC when Cray hired me and I wound up as head of communications and investor relations for the company.

Olds: Ah, okay. What kind of consulting projects were you doing with Cray back in the day?

Conway: Before I joined Cray as an employee, it was everything from employee surveys to talking to all of their investors and financial analysts and industry analysts to figure out how people saw the company so that the company, which was really growing fast at that point, could figure out how to steer well through this growth.

Olds: And doing, probably, quite a bit of education, too, because computers and supercomputers weren't that well known back then, were they?

Conway: Sometimes I needed to be an evangelist for HPC in an outward direction, but at the same time I was being educated by others in the company and in the HPC community.

Olds: So, what did you do after Cray?

Conway: Well, Cray got sold to SGI and when that happened, because of the position I was in, I was dealing with investor relations and all those things, the person in that job leaves. So, I knew I had to leave and I went into early Internet companies as an exec with a company called CompuServe, and then I was a consultant to AOL in its heyday.

Olds: Oh, I remember - I was a client of CompuServe's.

Conway: And then Cray became independent within SGI for a while, then independent totally, and I agreed to rejoin to help with the restart, especially helping the CEO to attract and maintain relationships with investors.

Olds: I remember those days. I was part of that transition from Cray to SGI, but I'm part of the piece that went to Sun.

Conway: I remember that. I worked very heavily with all the folks in Beaverton at Superservers, which was officially called the Cray Business Division. That division was sold to Sun and the system morphed

into the E10000 that was a major success for Sun. It was easily the fastest business computer on the planet then.

Olds: So, how did you get into the industry analyst game?

Conway: When Cray started up again I had my own consulting company, so I was kind of operating as an analyst and I told the new Cray that I would work there for two years and I worked there, to the day, exactly, for two years and then got hired by IDC as an industry analyst. IDC spun off its HPC team in 2017 and we became Hyperion Research.

Olds: You know at IDC and, of course, at Hyperion, you guys really have set the standard for what HPC market analysis and trend analysis should be.

Conway: Well, you know, there were a number of us, some of us went on to other things. Both Addison Snell and Chris Willard were colleagues and really helped grow the HPC business at IDC before they went off on their own. There are other people who are no longer with us but are doing a great job for the HPC community.

Olds: But it's quite an organization and methodology that you've built there. Over your years, what were some of the biggest changes that you've seen in HPC?

Conway: Well, I think the biggest change is the growth of the size of this market. In what some people remember as the "golden era" of supercomputing, up through the early 90's, when the competitors were Cray, IBM, Fujitsu, Hitachi, and NEC, the whole market with everything thrown in (hardware, software, servers, support, everything) was worth about \$2 billion worldwide. In 2019, it was \$28 billion, and we're forecasting it's going to go to \$43 billion by 2024, not counting maybe another \$7 or so billion in cloud usage and not reflecting the impact of the virus epidemic. So, this might be a \$50 or 51 billion market in a couple years. So, that's one big change.

Another big change is the systems themselves. They used to be these proprietary monoliths. They've become much more modular, standards-based and heterogeneous, and that's allowed the market to democratize and expand. Without that, I don't think it would have nearly as much. It's also evolved from kind of a niche, government-controlled market to a much more mature stage where it is a commercial market driven by commercial market forces and, for the vast majority of people, that's a good thing. Legacy stuff always gets left behind but, even for government, that's a much healthier thing. The other thing is the sizes of the systems have gotten so big and the prices as well. It used to be that \$30 million bought you the biggest supercomputer you could possibly buy. That new Fugaku system that's #1 on the top 500 that just came out is a billion-dollar contract all told.

Olds: Well, and it's probably going to cost more than \$30 million just to power the thing, or it's going to be a big number.

Conway: I think the other big thing that's happening here is that finally, after lots of work by lots of people, governments around the world have recognized that HPC is not just for science but it's crucial for economic competitiveness. That recognition is what has allowed all these huge sums to be made available for exascale systems. On the other hand, it's also led to the disruption of the market - the indigenous technology movements in China, Japan and Europe - because if it's so strategic the countries don't want to be reliant on foreign vendors, so there's this big push to develop indigenous supply chains and technologies. In the short term, that's disruptive; in the long term, if it goes right, then there ought to be more competition and choice in the market.

Olds: Exactly, and you also look at, and I think you alluded to this, HPC leaving the lab and going into the enterprise data center as well.

Conway: Absolutely. That started very slowly about ten years ago and it really, really picked up momentum. We track hundreds of companies that have adopted HPC for the first time. There is an important distinction here - we are not talking about the automotive and aerospace companies that are bringing HPC into HPC data centers for upstream R&D. This is for live, downstream, business operations like sales, marketing, HR, that kind of thing. They're adopting HPC because enterprise server technology by itself can't do it anymore in these big global companies. The enterprise server technology stays there in the data center but HPC gets integrated into it at very key decision points where very fast responses are needed to very complex questions.

Olds: Huge amount of data, very quick execution time because they are making decisions off of it in real time, almost. So, it is extraordinarily important to have the performance.

Conway: Yes, lots of companies we talk with used to produce reports from all their branches around the world once a month. Now they have to do it five or six times a day to remain competitive.

Olds: Yeah, it's incredible. You know what would be interesting? It's kind of spurred me from of your earlier remarks: if somebody ever computed out what the cost-to-the-customer per flop has done since, say, 1980 to now.

Conway: I think we, at least, used to do that. I don't know if we've done it exactly recently, but we have the ability to do that easily.

Olds: Because even with a \$1 billion system the cost of performance per unit must be way, way down. I mean, it must be dropping exponentially over the years, I would think.

Conway: It is. Part of that, of course, is the adoption of standard technologies for a lot of what happens in HPC. So that reduces the price, it expands the market because everybody can use the stuff and the programming is the same. But now that's sort of changing because, as you know, the market is now becoming, in a good way, much more heterogenous with the technologies, but also the problems that it's tackling are becoming more heterogeneous. And we're seeing the architectures, after a couple of decades where things became increasingly compute-intensive but not so great for data, we are seeing the newer architectures come out that are designed to support data-intensive work much more efficiently as well.

Olds: So, that's a good springboard to our next question: where do you see HPC going in the next few years?

Conway: Well, it's already going there. It's pushing into AI or AI is pushing into it. I'd say enterprise is pushing into HPC, at the level of the large global corporations, anyway. There's quantum computing. There's also the dance between on-premise and cloud computing that's happening. All of those boundaries are starting to melt down. It's kind of like, "resistance is futile." You really don't want to resist what's inevitable, though there's always some of that. But the workflows that these systems have to handle, the spectrum of types of work is growing enormously and that's challenging from a system design standpoint, challenging from a programming standpoint, but that's where it's going, like it or not.

Olds: I also see more, I believe, heterogeneity in systems with different kinds of accelerators, radically different kinds of accelerators in some cases, different designs. It was a big surprise to me to see that

Fugaku was all CPUs. No accelerators. But then I could probably show you a list of 50 or 60 companies, big and small, that are putting together their own accelerators.

Conway: Well, HPC is shifting to a workflow view which is more cloud-like. It's not the same as cloud microservices, but it's more cloud-like. So what we're seeing is architectures that are being designed to support, let's say, an on-premise workflow that might have to go through 20 different very lightweight containers. Each one of them has to very quickly assemble different hardware, software, and even storage resources. A workflow might go through 20 of these in a very short period of time. So, think about what we have always said in HPC, and sort of mean this jokingly, "hardware is easy, software is hard." So, the software continues to be the most challenging part of HPC.

Olds: Yes. So, what has you, if anything, concerned about the future of HPC?

Conway: Two things: one, that this indigenous movement around the world could get to be too nationalistic or regionalistic and collaboration could suffer as a result. I don't think that will happen, but in the short term I think there will be some of that. The other part that's, of course, concerning and particularly because I've been focusing a lot on AI is what Stephen Hawking expressed very well - that this could go in two different ways. That this could be a very, very good thing or it could be a not very good thing. I'm concerned that the ethical side of AI and all the rest of that will not be addressed sufficiently, that the momentum of the market will overtake those considerations. That could be a long-term problem.

Olds: That could be a very big problem. And it's hard to say how that problem is stopped once it gets going.

Conway: I think there have been some very good examples. Germany, in 2016, came out with the first national law for governing automated vehicles. They spent two years on that and they had religious people and ethicists involved. Who knows how it'll work in practice? We don't have fully automated vehicles yet, but that's kind of the way to go. I think China looked at what Germany did and is kind of headed in that direction. But we'll see.

Olds: So, final question: what has you most excited about HPC in the future?

Conway: What always has had me the most excited, and it's not the systems - they're tools like hammers and screwdrivers. It's what happens when this gear gets into the hands of some of the world's brightest, most creative people: scientists, engineers, data-analysts, and so forth. That's always the part that has excited me most. I'm paid to understand something about the systems, but what's exciting is really what people do with them.

Olds: And what comes out of that?

Conway: Yes.

Olds: Fantastic. Well, thank you so much for the time Steve, this has been a fascinating look at HPC and your roots and where you think things are going. Thank you.

Conway: Thank you, Dan. And I also want to give a big shout-out to my friend and former colleague Rich Brueckner.

Olds: Same here. Good job.

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

Copyright Notice

Copyright 2020 Hyperion Research LLC. Reproduction is forbidden unless authorized. All rights reserved. Visit www.HyperionResearch.com or www.hpcuserforum.com to learn more. Please contact 612.812.5798 and/or email info@hyperionres.com for information on reprints, additional copies, web rights, or quoting permission.