HYPERION RESEARCH

HPC User Forum Update

Interviews with Steering Committee Members: Ryan Quick, Providentia Worldwide

Steve Conway and Thomas Gerard
June 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 Ryan Quick. He is an expert at scale-out systems, UNIX kernel design and profiling, and has been recognized for innovation in hardware, application, marshaling, and high performance computing. His current efforts bring machine learning, real-time eventing, set-selection, and digital signal processing technologies to bear on predictive analytics and self-healing in command and control systems. A graduate of Vanderbilt University, Mr. Quick holds patents for messaging middleware systems, and is a pioneer in bridging High-Performance Computing technologies with enterprise best-practice infrastructure.

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.

RYAN QUICK INTERVIEWED BY DAN OLDS, insideHPC

Olds: Hello, I'm Dan Olds on behalf of Hyperion Research and insideHPC and we have a special treat today. We're going to be interviewing Ryan Quick. So, Ryan, can you introduce yourself a little bit?

Quick: Sure. My name is Ryan Quick, I've been with a boutique consulting firm called Providentia Worldwide for the last almost three years now. Before that I was with eBay/PayPal since early 2000, and we did what we called "atoms and galaxies" for them. We ran very large supercomputers for their fraud analytics and risk systems, and then we also did small, mini-board computers and custom hardware and things like that in the very small systems. Not the stuff in the middle. The very small atoms and the very small galaxies.

Olds: You left the middle alone.

Quick: We did that when we first got there in the early 2000s, but we left that stuff behind in about 2008 and went to the other stuff. So, for about a decade we ran their supers.

Olds: So, how did you get involved in HPC in the first place?

Quick: I had a problem where we had fractured their database, they call it now "sharding," we called it, "splitting" at the time. That worked really well, but you had to have the application, had to understand what you had done there. So, you could split it up to be able to make the database faster, but then you had to have your application understand what you had done in order to be able to put that back together. You call that the humpty-dumpty problem: you could throw humpty-dumpty off the wall, but your application had to know where you broke him up. That was great, but then your analytics people were like, "well, we don't know how to put him back together again." Hence the humpty-dumpty problem.

We had been tasked with constructing something like a data warehouse to put it back together again. So, we had this problem where we could either keep the database up-to-date by putting it back together again or we could query from it, but we couldn't do both at the same time. So, we could build a single database that was fast enough to keep all the updates coming in, but we had to stop it if we wanted to ask it a question. That wasn't going to work. So, I went shopping on the Internet and was like, hey, in my before time I lived in northern Virginia and we used to run, in my basement, I had a Challenge XL and some other gear I had picked up from NASA in the sales that they used to run out at Oceana Naval Air Station and places like that.

We would pick up stuff that we would use as heaters that happened to have reality engines and stuff like that in them. I was like, "I haven't used a real SGI in a long time. I wonder if they've got something that might be able to do it." I went shopping right before Christmas, this was 2004, right after Thanksgiving I cold-called SGI and said, "hey, I'm looking at this Altix," the 450 was what I wanted, and said, "hey, you don't really have anything on there that says that you can do this, but I'd like to buy one of these things, a baby one, and I'm going to try and run my Sequel on it and see if I can build a database that can simultaneously keep this update rate running and be able to run queries on it." And that's what we did. I went shopping, bought an SGI, and we actually were able to do that. So that's what we did and how I got involved in HPC.

Olds: They must've been pretty interested in that use case.

Quick: They were. Well, first I had to convince Chippewa Falls that running a database on top of an Altix was something that I was allowed to do. And we quickly moved off the Altix slightly before it became UV [SGI UltraViolet]. It became UV right after that. We moved off the Altix, we kept that one around, because I did buy it, and then we moved to the [SGI] ICE platform after that. Then, after that, we kept a hybrid, because we actually had use cases that were good for the large, shared memory footprint and for the scale-out platform. So, we always ran a hybrid cluster with a mix of both. We had both UV and scale-out ICE from then on out.

Olds: So, what are some of the biggest changes you've seen in HPC over your career?

Quick: What HPC means to different people. It used to be taboo, for wherever you were, to talk about anything different. So, if you were a weather guy, HPC was weather; if you were a physics guy, it was physics. Now, it's gotten to be a little bit better. You can talk about not just what you do, but you can talk about what other people do. In fact, I was asked by the SC19 committee last year in the summer to write an article for them, so I wrote a blog piece back last year specifically about that. I called it "Open HPC," where basically the theme of it was, we've finally gotten to where we can actually talk about different types of HPC. Wouldn't it be nice now if we would actually sit in each other's BoF [Birds of a Feather] sessions and maybe start sharing what it means to us? Maybe we can learn from each other and not just getting elbowed in the face as we walk along for coming from different areas. We've stopped doing that. Maybe now we can learn from each other. So, yeah, that's probably the biggest thing, is not only do you not get in trouble for having a different pedigree, now we are getting to where we might actually be able to learn something from each other. That's what I'm hoping.

Olds: That's a heady world that you're imagining. Looking forward, what do you see that either excites you or concerns you?

Quick: Well, let's see, "concerns me" is probably the easiest one to start with and that is, that we will close the doors on innovation in the name of speed. And that actually comes because of the COVID stuff. I think COVID is a horrible, unfortunate thing to happen to the world, obviously, but it's actually a big boon to large-scale computation in the world. Just like it's speeding up how we're going about doing drug trials and things like that, it's also going to speed up how we learn about large-scale computing, because we are going to finally try some stuff we've never tried before. But what I'm also afraid of, at the same time, is that we're going to get one new thing that works and that's all we're going to do. We're going to figure out one new way of something and then everybody's going to be like, "well, great, we learned something new!" And then they're going to slam the door behind them. So, rather than keep the screen-door open and allow a few more things in and try several things, we're just going to get one new thing that works and go with that. That's what I'm afraid of. So, that's probably the biggest thing.

Olds: So, whatever methodology, for example, solves the COVID-19 crisis that'll be the one Holy Grail and we won't pursue anything else.

Quick: Yeah, it'll be like CUDA 2. There could've been a CUDA 3, or CUDA 4 or a something-other-than-CUDA. So, that's probably the thing I'm worried about. It's software, right? There's lots of ways of solving software.

Olds: Sure. What excites you?

Quick: Yeah, the thing that excites me most is that, finally, I've gone six months now and I haven't been asked to explain why containerization is a different way of doing virtual machines, which is great.

I'm not even going to go into it. I'm going to let that stand. Now you're going to have to explain it in the comments. I'm just going to let that stand. But that's good, because I've made it six months without having to justify that statement, but that means that we've finally gotten somewhere with containerization and that means it's come into its own in the HPC world, which should open the door for all of the things that containerization could actually do for us and not simply look at it as a way of lifting and shifting workloads, because that has been in our way for over a decade now and I am finally going to get to see containers come into their own.

So, a little bit of history there for me. In 2006, I was working for PayPal at the time, we went before the Visa committee to get Solaris containers approved for the PCI committee to allow Solaris containers to be approved for doing payment card industry isolations for Oracle. Specifically, Oracle database running inside of Sun Solaris' (this is prior to the acquisition). So, I've been sitting here like, look, if Oracle databases were allowed to process payment card in isolation inside containers in 2006, why are we still debating this technology in 2020 as being isolation tech? For me it's like I can't believe I'm still having this discussion. So, it's finally getting there with Linux and that's making me really happy, so that's good.

Olds: Very nice, the fulfillment of a dream. Where do you see HPC headed in the future? Are there any other trends that you are looking at, keeping your eye on?

Quick: My company now, we work in this gray area where hyperscale, AI and HPC all sort of meet and mix, because there is this sort of Prometheus-bringing-fire across the way in between those, and that's kind of what we do is sort of beg and borrow technologies between the different entities and bring them to the other side—mostly because the vendors aren't well aligned at being able to bring technologies from one business unit to the other side.

So, I won't name names here, but pretty much all of the different vendors have technologies that are capable of delivering solutions across the stack. But usually one particularly technology is aligned with a particular core business unit. So, high-end clustering might be associated with the HPC business unit, whereas scale-out object technology might be associated with a different type of storage in the hyperscale unit. Well, if you try to take clustering and wanted object stores together, they may never talk. So, trying to buy that might be nearly impossible, depending on who your sales reps are. And unless you're a tier one enterprise where you can force the right people to be in the room for you, you may never be able to buy that from the company, even if you can go to the website and say, "Look, right there's the SKU, I want to buy that."

So, I think we'll finally, in the HPC world, start to see some of those technologies really start to mix and match and that'll be a good thing. And I think it'll be driven by edge and IoT, especially, because they're starting to take some of the really high-end HPC-class hardware and cram it into smaller stuff, which is great because now you're going to see small stuff start to do some really, really neat things. It's one thing to have a cool drone flying around your house, but what if that drone is processing 16 or 24 4K video streams at one time and wants to be able to do facial recognition as it flies around your neighborhood? Now that's some cool HPC that's happening while that thing's flying around and something's got to analyze that stuff.

Olds: That's some serious back end. Well, fantastic. Thank you so much for the time, Ryan. We really appreciate it and I know that all the listeners and watchers out there are going to vastly enjoy this. Thanks again.

Quick: Yeah, I appreciate it. Thanks, Dan.

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
