Hector Garcia-Molina Speaks Out

Regarding Startups, How Life is Getting Harder, Delta Papers, CS in Mexico, Life as a Department Chair, and More

by Marianne Winslett

Hector Garcia-Molina

http://www-db.stanford.edu/people/hector.html

Welcome to this installment in the *SIGMOD Record*'s series of interviews with pillars of the database community. This issue's interview with Hector Garcia-Molina took place in June 2001 (face to face) and October 2001 (email). Hector is the Leonard Bosack and Sandra Lerner Professor in the Departments of Computer Science and Electrical Engineering at Stanford University, and is the chair of the Department of Computer Science. Before coming to Stanford, he was on the faculty at Princeton University. He is a Fellow of the ACM, received the 1999 ACM SIGMOD Innovations Award, is a member of the President's Information Technology Advisory Committee (PITAC), and is a member of the Oracle Board of Directors.

"[W]hen I look at what companies are doing---and especially startups---they rarely have the time to do anything really interesting and significant"

During the videotaping of the interview, we ran out of time before we ran out of interesting topics. We decided to continue the interview later on, with the result that the portion of the interview above the ruled line was conducted face to face, and the portion below the ruled line was conducted via email.

The usual caveats: To let the conversation flow freely and easily, I videotaped the interview and transcribed it later. The errors of transcription, the changes necessary when speech is converted to the written word, and any edits for length or ordering are my own. Eventually the videos will appear on the SIGMOD web site, so you can directly hear

what each pillar has to say, without my editing and with all the original nuances of emphasis that are eliminated during transcription.

In upcoming issues, we'll be hearing from Phil Bernstein, Jim Gray, David Maier, and Mike Stonebraker.

What do you think is the most exciting trend in technology today?

One thing that interests me---I don't know how significant it is---is peer to peer systems. I've always been interested in distributed computing [issues]. Even though I've been working in the database community, a lot of my work was with distributed systems---when you have data in different places, or copies of data on different computers, and how to deal with [those data], synchronize them, coordinate them.

What do you mean when you say "peer to peer"?

"[D]on't worry too much about tenure"

"Peer to peer" is a vague term that people are using in different ways. What I understand [by "peer to peer"] is that you have a set of computers that have [roughly] equivalent capabilities. There is no specialized node that is controlling the system, or is holding a central index, for example, that everybody can consult. Of course centralized systems are good for many things, but having a system where everybody is [roughly] equivalent and can work together has some real advantages---plus it's extremely fun to work with those systems, because it's very challenging. How do you find resources, for example, in such a system, where you can't just go to the, [say,] Napster index, and [say], "tell me who has this song." You have to figure [out who has the song] by talking to your peers [to] understand who might have [it]. So the problems are much harder, but that's why it's fun to work on them.

I've heard it said that all the really exciting database research is taking place in industry today. What do you think about that?

I disagree with [that]. There is [some very] interesting work being done in industry, but when I look at what companies are doing---and especially startups---they rarely have the time to do anything really interesting and significant. They always take the simplest solution to the problem, try to get it out as quickly as possible. They never take the time to really understand why they did something one way, or if they could have done it in another way. If they can get it out the door, that's all they care [about]. And I don't think that's as interesting or fun or as valuable as what we do in universities. [Of course] some companies, some research labs, [still] take the time to do this [careful analysis]. But without people taking the time to understand … even if we didn't come up with new ideas in universities, even if we just take what's out there [from industry] but analyze it, study it, organize the ideas so that people can learn and improve things, I think that that's a great contribution. It *is* a great contribution, I think. And people tend to forget that side

of research when they talk about companies doing research. [Most startups] don't do real research. They come up with some things and quickly implement, hack things together, but they're not really doing scientific research.

Is that a problem, or is it okay that we have that separation [of function]?

I don't think it's a problem. It's not that separate, we do have a lot of interaction with companies. They come and see what we're doing, we go and talk to them. So I think it's fine. The only thing I don't agree with is this notion that's become popular, that all the fun research in databases or in computer science is being done in companies.

"When I was graduating, having one or two or three publications was wonderful. Now if you don't have ten publications, people say, 'what's wrong with you?'"

What problems should people be working on today that they aren't?

Well, one of the more important ones that we're not spending enough time on is user interfaces, or user-related problems. How do we make database technology more accessible, easier to use? How can you visualize what's in the database in a simpler way? How can you manipulate the data without having to be a database expert? There's a lot of work that has to be done. [I] know people complain that it's hard to get papers published, in the database community at least, [on user-related topics]. And it's true [that it's hard to get papers published on these topics. When] dealing with users and user interfaces, it's hard to quantify what the contribution or what the benefit of your technique is. [On the other hand,] if you are dealing with performance, you can come up with an algorithm, and say, okay, I can process ten queries a second [using my algorithm] as opposed to five queries a second [without it], so [my algorithm has] done something. But with user interface issues, it's harder to quantify [the contributions there. Perhaps you've] saved people time, but how do you measure that?

Is [user-related research] something that can be done more easily in industry, where they are closer to users in a certain sense?

Again, I don't think industry really does very much research. They come up with an idea and they try to sell it. If it was a good idea, maybe they'll make money. Even if it was a bad idea, if they have good marketing people, they might still make money and we never know: was the product successful because they just have good marketing people and could intimidate the competition well enough, or was it really a good, useful idea that they had? I think people in industry have good ideas in the [human-computer interface] domain, but I don't think they have an advantage over [academics] in testing the ideas and evaluating them and performing measurements and really understanding what are the right techniques.

When you say that, it makes me think back to Mosaic, the first popular web browser. In a sense, that was just a triumph of user interface [engineering]. There was no research [involved in its creation].

"I call [the way we do CS research] the "slash and burn" approach to science"

Yes, and that was why it was successful over, for example, the [browser] called Gopher. [Mosaic] had a much nicer user interface.

Being a professor today is a serious juggling act between work, family, and outside commitments. You went through the tenure system at Princeton rather than at Stanford, but I've heard it said of Stanford (many years ago when I was here) that those who got tenure also got a divorce. What recommendations do you have for people to handle the juggling act?

I guess I'm in the category of people who got tenure at Princeton and got a divorce.

Oh, I didn't know that.

It wasn't exactly at the same time, but soon after.

I guess that makes you an expert. What tips do you have for people who are going through this process?

Going through a divorce, or going through tenure?

The tenure process!

It's really hard [to give advice … The] advice I give to junior faculty [who] I'm supposed to be mentoring in our department [is:] don't worry too much about tenure. [Of course,] you have to do certain things to try to get tenure. But decide what to work on mainly [based] on whether you like the topic, whether you are enjoying [working on] it. [If] you just do everything based on [whether it] will get you tenure or not, it can be very frustrating. I think, if you are doing work you enjoy, whether you get tenure [is] not going to be that critical. If you don't get tenure, you'll go somewhere else and do fun work.

Why don't I know more computer scientists from Mexico? I seem to know more Canadians than Mexicans.

There's no research tradition in Mexico. When you go to school there, people are more interested in making money. I think it's common in a lot of countries, where you get prestige and are successful if you make money, not if you teach at a university or do research. [At least when I was a student in Mexico, there was no research tradition, although things may be starting to change now.]

[Then how did you end up in a research career]?

Partly because my family did have this tradition. My father was a university professor, one of the few who appreciated that [line of work]. I got it from [my family].

What do you know now that you wish you'd known as a fledgling computer scientist? And any other words of advice for those just starting out?

It seems that things have gotten harder, in terms of publications. When I was graduating, having one or two or three publications was wonderful. Now if you don't have ten publications, people say, "what's wrong with you?" I'm only exaggerating a little bit, not ten but---

I understand; I've seen it too.

I think there's a lot more pressure on [graduate] students right now, both because they have to publish more, [and also] because there's so much activity in industry on related [topics] that there's a lot of pressure to differentiate yourself. Before, I think, we were just sort of doing things that we thought were neat and fun, and working on [those fun topics], and if you got a paper out of your thesis that was fine, and if not, that was fine too. [It] seems that it's tougher.

"[P]eople talk about how much work it is [to be] an assistant professor and trying to get tenure. And you somehow, directly or indirectly, get the impression that oh, once I get tenure, things will improve and life will be less hectic. But [that] is actually not the case …"

Is that good or bad?

I don't know; it's just tougher.

Do you have any words of advice for mid-career database researchers or practitioners?

Before you get tenure, people talk about how much work it is [to be] an assistant professor and trying to get tenure. And you somehow, directly or indirectly, get the impression that oh, once I get tenure, things will improve and life will be less hectic.
But [that] is actually not the case, because once you get tenure, [they] start putting you on committees, asking you to do more things, and people expect more of you, so [the hectic nature of your work life] just continues and maybe gets worse a little bit.

Do you think we should really let [the assistant professors] know that, though? The assistant professors may all quit today if we tell them.

That's right, we shouldn't [tell them], we shouldn't.

Well it's the same way with a grad student just finishing their thesis. They have no idea that [their work life] can actually be harder than it is right then for them.

How do you keep your research program going while you are the chairman of the Computer Science Department at Stanford?

Well, we have an excellent department staff that takes care of a lot of day-to-day issues. So being chair only takes about two or three days a week, sometimes even less. This leaves me a full four days for my research… My students may not like it because I often meet with them during the weekend… But I have great PhD students who do a lot of good work on their own, so it all seems to be working out! Plus, my appointment is only for three years, so hopefully my brain will not deteriorate completely with bureaucratic nonsense by the time I am done!

I know you are on PITAC, and have recently been appointed to the Board of Directors of Oracle. What do these jobs involve?

PITAC is the President's Information Technology Advisory Committee. There are about 20 of us in this committee, from academia and from industry. (Jim Gray is the other database person on PITAC.) We are charged with giving advice to Congress and the President on information technology, especially on the research side. Three years ago we wrote a report that basically said that IT research funding should be doubled in five years, from about one billion dollars to about two. The first two years there were indeed significant funding increases, especially at NSF. The increased NSF funds lead to the ITR program, among other things. The federal research funding picture may not be as positive in the future, but our committee continues to work on making the case that IT research and development is critical for the nation, and overseeing how the allocated IT funds are being spent.

"What could be argued is that the whole database community has changed, now valuing more applied work."

Being on PITAC has been a great experience. I have learned a lot about how our government operates. There are a lot of excellent people in government, but they need all the support our community can provide, for example, [in] making the case to the general public that science and technology are important for the country. Also, it is critical for members of our community to participate in government activities, [such as] being on advisory committees, or working at NSF, DARPA, NASA, DOE, etc. as a program manager for a couple of years.

At Oracle, I just joined their Board of Directors. I think this is an exciting opportunity for me, especially since Oracle is one of the leading and dominant companies in our database field. (Perhaps now that I am [on] their board, I should say *the* leading company!)

Hopefully, I can help as a bridge between our research community and Oracle, of course, with "traffic" on the "bridge" going in both directions.

You did your graduate work at Stanford, then went to Princeton, then came back to Stanford. How was it to come "home" after having been at Stanford as a student?

It definitely felt strange for a while. When I arrived at Stanford I was assigned my former advisor's office (Gio Wiederhold had gone to DARPA for three years), so it felt strange to be in that office, but on the "wrong" side of the desk! On the other hand, I had been away long enough, that no one treated me as a graduate student. My current students now joke that some day they will return to Stanford as faculty and will kick me out of my office. (pause) Well, I think they are joking…

Some people in the database community have said that the work coming out of Stanford is not as fundamental as it used to be, that it's more in the nature of delta work, that Stanford is more of an incubator for startups than a place for basic research, that the students are as bright as ever but the thoroughness and quality of their work has decreased, perhaps as a sign of the times. Can you comment on this?

Were these people at Wisconsin by any chance? Just kidding, as you know we have a friendly "rivalry" with our good friends in Madison. Anyway, I do *not* think Stanford has changed relative to other departments. For example, if you look at the publications on our web site, you'll see that we continue to publish at a decent rate, in the same places we used to publish. What could be argued is that the whole database community has changed, now valuing more applied work. I think this may be true; for instance, I recently got a review from a top database conference blasting our work because it was not "commercially feasible."

"My students may not like [my being department chair] because I often meet with them during the weekend"

I have also heard people complaint about the popularity of "delta work" in the database community. I do not think our community is any different from other computer science communities, and I do not see a big change in the number of delta papers over the years. Part of the problem is that our field is maturing, and we just do not have as many real "breakthrough" papers. Maybe now that we are more mature, we may even begin to follow the more scientific approach, where someone publishes some results, and then other people try to replicate the results? As you know, in our field, no one could ever publish a paper that just says, "Hey, we are re-running the experiments that so and so did, just to check if they are true, or if we can do better." Today, the referees would give you the lowest possible score: It's been done, so who cares??!! I call this the "slash and burn" approach to science. Once I publish my idea, no one can study that same idea, everything has to be new. Hopefully, this approach to our science will change in the future.

"[P]eople in government [need] all the support our community can provide, for example, [in] making the case to the general public that science and technology are important for the country."

I heard that when one of the major database conferences switched to double-blind reviewing, the number of acceptances of papers from some well-known database research groups went way down. Is this a good thing?

I have also heard the opposite, that some other conferences that tried blind reviewing did not notice any change, and hence dropped it. I personally have not noticed anything special yet. I personally do not like blind reviewing, one, because people can usually tell where a paper is from, even without the names. And two, as a reviewer, it tells me the conference organizers do not trust my judgment.

Your TSIMMIS project targeted information integration. What still needs to be accomplished in this area? Are there still barriers to the effective use of information integration in the intelligence community?

There are definitely still many hard, open problems in information integration. I believe that Tsimmis and other integration projects at the time, like the Information Manifold, Garlic, etc., made real contributions, but now the bar keeps getting higher. For example, now people would like to take into account more and more of the data semantics. With the recent events of September 11, I see a bigger role for information integration, and it becomes critical to combine information from diverse sources, in order to "make connections" that could not be made within a single database.