Kian-Lee Tan Speaks Out on How to Build a Strong DB Group without Pushing Students Hard

Marianne Winslett and Vanessa Braganholo

Kian-Lee Tan

http://www.comp.nus.edu.sg/~tankl/

Welcome to ACM SIGMOD Record's series of interviews with distinguished members of the database community. I'm Marianne Winslett, and today[1] *we are at the Advanced Digital Sciences Center in Singapore. I have here with me Kian-Lee Tan, who's a Provost's Chair Professor of Computer Science at the National University of Singapore, where he's also the Vice Dean for Research in the School of Computing. Kian-Lee received the President's Science Award in 2011. He has served as co-editor in chief of the VLDB Journal and editor for IEEE Transactions on Knowledge and Data Engineering, and a PC chair for VLDB and ICDE. His PhD is from the National University of Singapore. So Kian-Lee, welcome!*

[1] This interview was conducted in 2011.

So why did you receive the President's Science Award?

Actually, the PSA is an annual award. It is Singapore's most prestigious award given to researchers in recognition of their outstanding contributions in basic research. It is actually a team award, rather than my award. For 2011, Beng Chin and I have the honor of receiving this award for our research on Peer-to-Peer Based Data Management. And, of course, thanks to you for nominating us!

You're welcome! So, that's a great lead into my next question. Which is, so far, peer-to-peer databases haven't had a big impact in industry. Do you see a killer app for them?

Actually, no. To be frank, we were quite disappointed that we were not successful in bringing this technology into industry. We have tried talking to some people, particularly the clinics in Singapore, but I guess they are more concerned with security. So we were not very successful in bringing this technology out. But nevertheless, Beng Chin actually founded a company called BestPeer, that is making use of this technology and he seems to be making some headway in China. So, exactly what is being done is being kept tight-lipped, since it has to do with some government agencies in China.

So not health care?

Right, not health care.

I see…

So there's really no "killer app", but there will be organizations that are interested to look at it.

Mike Stonebraker says that MapReduce is a major step backwards. Do you agree with him?

I don't think so. I mean, you have to look at the objective of MapReduce, right? It's not intended for database applications. The fact is that many people are using MapReduce, especially in the industry, so it is quite a successful framework. From that perspective, I suppose he's referring to database management systems. He's comparing to database management systems. So in that case, I don't think it's right.

Okay, wrong comparison. So, nowadays companies like Microsoft and Google have lots of money, they have hundreds of researchers with PhDs and they have all the data that anybody could want. So how can we as academic researchers compete with those places?

Don't throw everything into a single basket.

First, I think there are lots of problems to be solved, beyond what Google and Microsoft are doing. Of course, they have the benefit of real life data. For us, perhaps we need to wake up and work on some real applications. For example, in NUS now, we have a number of big grants that allow us to build test beds. For example, one of them is on megacities. The government has agreed to set aside some area for us to build our sensors both in China as well as Singapore. So, it's a joint project between China and Singapore. With that, we'll be collecting real data and hopefully we'll be able to do something that is both relevant and impactful.

What kind of data will you collect?

This particular project has to do with air pollution and water pollution. In some sense you can look at it as streaming data. They'll be sensing the pollutants in the air, in the water, and then this will be transmitted to the central live server for real-time processing.

That's an interesting application. So what are the research challenges that you see so far in that?

I think the main thing for us is, we need to understand the domain. And then after that, we need to ensure that this information is collected in real-time. Results need to be presented in real time. So in our part…the contribution, rather, the part that SoC plays is to build a backend server that will be able to process this streaming data quickly and provide information to those who need them in real time.

Okay. Great! So you've been in NUS ever since you were a teenager: Undergrad, Masters, PhD and then on to being a professor! So at that point in time, NUS wasn't known for its research, when you got your PhD. So many of your peers probably went overseas for their PhDs. So how did staying at NUS affect you? In your career?

I don't think it affected much, in terms of recognition. I guess the reason is because we started off well. I mean, my advisor began by publishing in either ICDE or VLDB. So, there's already a certain bar that he has set. And when Beng Chin joined, he also encouraged

us to submit to all these prestigious conferences and journals. So from that perspective I think people know our work and know us. Personally, I think, at the end of the day, nobody remembers or nobody cares where you get your PhD from. I probably don't know where you received your PhD. You may not know where (I received my PhD) until now. I think people recognize your work rather than where you received your PhD.

So the database group at NUS is very well known today and you played a key role in building that up. Can you share your insights in how to make that happen?

Actually, I'm not the only one to take on a key role. I think it's a team effort. I mean, there are other colleagues in the DB group. There are a few things that I think we have done. One is, we actually rank conferences. That means, SIGMOD, VLDB, ICDE are considered Tier 1. And the other conferences are Tier 2, Tier 3. In this way, we effectively encourage our people to submit to Tier 1 first. So everybody, including our students now, whenever it comes to the deadline for SIGMOD, ICDE, VLDB, everyone wants to submit something there.

And then, before you ranked them, what were people doing? So how did that change their behavior? What were they picking to send their work too?

I suppose it was whatever deadlines were available. So now, people may hold back.

Interesting! Did you do other things too?

Yes, I think another thing we have done is established collaboration with overseas researchers. We send our students over for internships. Given that these are better students, good students, to some extent, it helps to establish our credibility. I think we try to increase our visibility by being more professionally active.

Do you mean like in the societies and the journals? Or organizing conferences? Or?

Right, as PC members and things like that. In terms of recruitment, nowadays it is not so much on paper publications that we are looking at.

Oh? What do you look at?

Well, I suppose the impact of the work. I think we have published well, and we would like to have impact beyond just paper publications.

So for a young person, how do you judge the impact?

I guess by the quality of their work. We assess what they have done for their PhD and whether it is a strategic direction that we are interested in. In particular, we are very interested in recruiting those who build systems, so we are more systems oriented now.

Now, you don't like to travel. I'm told you don't even go to conferences.

That's wrong! I attend one conference a year.

Oh okay, well is it a conference in Singapore? Or a conference far away?

Not far away. Usually I attend once a year. And for the next six years, I'll be attending VLDB, because I'm the EiC for the VLDB Journal. Recently I just joined the VLDB Endowment. So there's a reason for me to travel. I try to minimize traveling.

Okay, well then how come DBLP says you have 182 co-authors if you're not traveling?

I actually looked through the list. I think most of them are students at NUS.

That many?

Quite a large number. For example, I work with Beng Chin, so Beng Chin's students are in the list. I work with Anthony Tung, and Anthony's students are in the list. Some of them are interns from Chinese universities. So there's quite a lot of interns that come from Chinese Universities. In addition, I've been around for 20 years.

I've been told you catch your bus at 5pm every day. Even if there is a deadline. And that you're also active with your church and with your family (you have 2 kids). So how do you manage to balance all these things?

First, I actually arrive at work at 7am or so, right? Otherwise my boss would be upset with me. Yeah, the reason for leaving early is to beat the traffic. Otherwise my traveling time would be very long. So I always do it to hopefully always get a seat in the bus.

Okay so how long is your traveling time?

At non-peek hours, it is probably 45 min, and at peek hours it could be as long as 1.5 hours.

So you must leave, if you get there at 7am, you must catch the bus at 6:15am…

Yes, that is right.

6:15am?!

Yes.

(laughter from both)

And then, it's still a 45-minute ride. Well coming from Champaign, 45 minutes to me is practically forever. Actually 45 minutes, that's all the way across Singapore, right?

Yeah.

How do stand such a long ride?

Actually I find my journey to be quite productive. Usually I read in the bus. I can review a paper. I read books. So I think it's not wasted time. All I need is a seat in the bus.

Okay, and is it the same when you go home at 5?

Right, that's right.

So they're not too crowded at that point. So really, your workday is 6:15am to 5:45pm and that is a pretty long day.

Yup, I'm glad to hear that.

Wait, unless some people work late at night also.

Yeah, I think the students are expected to work from 10 to 10. But that doesn't mean I stop to work at 5PM.

We need to wake up and work on some real applications.

Then how do you fill the other stuff? The church thing and family time…?

So, I keep my Saturdays free for sure. I try not to be involved in any other activities, except with my family. If possible, no church activities, no office activities. I spend my Sundays at church, from morning until 6PM. So, I have five full-time days related to office work. If there is a need to, I may have to review a paper at home after 5PM. In fact, for the last SIGMOD, I worked until 3AM.

Woah! Okay. And what about your kids? If they're in high school they're working until midnight themselves, aren't they?

No, my son is going for the University next year.

Oh okay!

My daughter is coming to Secondary 2, which is 14 years old.

I've been told that you never push your students hard. But some people think they have to push their students hard. So, how does that work for you?

Is that a compliment? Or….? (laughing)

A complaint? No no! I promise the person who suggested I asked this thought it was great, so they weren't complaining.

I think it's a matter of style. For myself when I was pursuing my PhD, my advisor did not push me. So I was pretty much independent, but I saw my advisor regularly. So in some sense, I expect the same thing of my students. I make it clear to them that they should be independent, that they have only 4 years of scholarship. To some extent, that is indirectly pushing them. So if they don't produce within 4 years, if they can't graduate within 4 years, they might not have financial support beyond that. So far, it works. So I don't like to scold people.

So if you were a scolder does that mean… would your DBLP… would you produce more? Or do you think it's a matter of style and it doesn't directly have any impact?

I think the students already have a lot of pressure, especially from the DB group. I mean, there are quite a number of students who publish in the so-called Tier 1 conferences and journals, and those who have not are really pushing themselves hard.

So you've had a very successful career. What's been your biggest failure?

At one stage, I'm interested in exploring bioinformatics. I actually graduated two PhD students in bioinformatics research. But I view this is still as a failure in the sense that what I've done is still more from the CS point of view, rather than contributing more towards the science. So actually, I have a joint appointment at Genome Institute of Singapore. And

I think people recognize your work rather than where you received your PhD.

after two years, I have not been able to interact well with people there. I guess partly it's my fault, in the sense that I didn't really interact well with them. Because in the initial stage, the scientists there were more interested in getting the CS people to develop tools for them. They are doing manual stuff and they want to automate it. So from my point of view, these are more engineering. So in the end, it did not work out. I wish that more could have been done.

So probably some of our readers are trying to collaborate with biologists, so what advice do you have for them? What's the right way to connect with the biologists? How can you make it work?

I think you must be prepared to spend more time with them, to familiarize with their domain. Perhaps to even be prepared to do some engineering work for them in the process.

I hear that you're a collector. What do you collect?

I collect Coca-Cola cans and bottles from different countries.

How many do you have?

Oh I didn't count, but usually when I go to a new place, if I find a can that I don't have, I'll buy it and bring it back. And now that my students are aware of it, sometimes when they travel, they'll bring one for me too.

I see. Because I was going to say if you don't travel, your collection would be very small! Maybe twelve cans?

No, I have more than that. My students will bring them back for me.

I see… okay. So why Coke cans?

Why Coke cans?

Are you a Coke drinker?

Not really, I drink coffee.

(laughter from both)

So why Coke cans… I supposed it is just to be different from others.

Okay… it is different! Why do you like research?

You get to do what you like. You get to develop new knowledge, create new knowledge that will be hopefully useful to society. We have that flexibility to do whatever we want.

Okay good! Why do you keep your clocks 15 minutes ahead?

So I can leave at 4:45 now (laughs). No, I'm very particular about punctuality. I don't like to be late for meetings. I think it is a mark of respect to others. I don't like others to wait for me, and I also don't like to wait for others. Even with my wife, when we go shopping and if I arrange to meet her at some place after a certain time -- if she's late, I'll be upset too.

But you'll think she's late because your watch is 15 minutes ahead!

No no! I will know how to adjust.

But somehow it helps keep you on time?

Right.

I have been told that the emails you send are short enough to be a tweet. "See me –KL". "How? –KL". "Status? –KL". So why are your emails so short?

Usually I actually write something longer, but in the end I always shorten my emails after writing it. I think in this case, there is a context behind it, which the recipient will know. I suppose this comes from my students, so we must have discussed something before and so I expect a response. So they would know what I'm talking about.

And their answer is very long?

Could be, but usually I ask them to see me rather than looking through their response.

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

I would say "don't throw everything into a single basket". Don't just work on one single problem. It would be too focused. Sometimes it's good to explore other fields at the same time and other areas at the same time. I mean, this is what I tell my PhD students. They should do more than what needs to be included in

their thesis. I guess it's the same as my own experience. For my PhD thesis, for example, I did more work than what is included in my thesis. I think that will give them a broader view of what is happening in that area, rather than being too narrow.

I think that's very good advice because my PhD advisor only gave me two pieces of advice and that was one of them. Although his was a simpler version, he said, "Always do two things".

Okay, among all your past research, do you have a favorite piece of work?

I think my favorite earlier work would be my PhD work on query optimization in parallel databases. But if you talk about my more recent work, our work on query authentication is something that I like. I think one aspect that has still not received attention is the concept of minimality. What it means is that "only answers are returned". Most of the current work requires exposing certain boundary points. So I think ours is the only work that doesn't require exposing boundary points.

I see, so you came up with a way so that they can verify the answer is complete without showing some things that are not part of the answer. And so, why don't you like the "show the boundary" approach?

So effectively, it is releasing more information than is required for the user.

So it's not exactly Access Control, but yeah, I understand.

If you magically had the time to do one additional thing at work that you're not doing now, what would it be?

I actually wanted to write a book every three years.

Really?!? Wow.

Right from the beginning, I actually co-authored a book every three years. After that, I stopped. So I think if I had time I would be …

What would the topic be if you write it yourself?

I think at this moment, I am interested in security. Database security. I'm teaching a course on security, and there is really no good textbook out there. So it is something I would be interested to consider.

If you could change one thing about yourself as a computer science researcher, what would it be?

To be more pro-active in talking to other researchers that are not in database research. In that way, more interesting research ideas could come out of it. For example, now, we're doing some work on PCM (Phase-change-memory). If I could spend some time just talking to colleagues that are working in that field, maybe our idea could have been more interesting.

So what's the relationship between phase-change memory and databases?

I suppose it's the technology, right? Phase-change-memory allows you to keep more data in memory, while still facilitating the random access.

Thank you very much for talking with me today.

Thank you Marianne.