Beng Chin Ooi Speaks Out on Building a Strong Database Group

Marianne Winslett and Vanessa Braganholo

Beng Chin Ooi
http://www.comp.nus.edu.sg/~ooibc/

Welcome to ACM SIGMOD Record's series of interviews with distinguished members of the database community. I'm Marianne Winslett, and today we are at my office at the Advanced Digital Sciences Center in Singapore, an outpost of the University of Illinois. I have here with me today Beng Chin Ooi who is the dean of the school of computing at the National University of Singapore where he's been a professor of computer science for many years. Beng Chin is editor-in-chief for IEEE Transactions on Knowledge and Data Engineering. He is the recipient of the 2009 SIGMOD Contributions Award, and he is an IEEE and ACM Fellow and Fellow of Singapore National Academy of Science. He is the co-founder of two startups and his Ph.D. is from Monash University. So Beng Chin, welcome! **(Please note that this interview took place in 2011)**

Okay, so, what's it like to be the dean?

That's an easy job. It's great! As a dean, I attend two morning meetings with the bosses and do lunch meetings with heads, vice deans, and assistant deans. So that means I have four meetings to attend a month, and the job is easy.

I don't believe you, but we're going to come back to that topic later. Don't believe him, readers!

OK, so let's switch to something technical. Cloud computing is hot. What will be the next big research issue in cloud computing?

For cloud computing, since it's a distributed cluster computing environment, issues such as data and system security, transaction management, efficiency, and query processing strategies are not yet fully resolved. They will be solved eventually, and of course, there are many other challenges that will come along as we make advancements in the hardware. For example, suppose the network bandwidth increases... then we will have different problems to solve. Also, with the introduction of the PCM (Phase Change Memory) chip and all other things, we will have to redesign architectures...

So why is transaction processing performance lower in the cloud than in a non-cloud environment?

In the cloud environment, because it's a distributed environment, suppose you want to enforce ACID properties. Then the locking overhead is going to take much longer, and therefore people go for less consistency for the sake of performance.

So what's the new hardware that you mentioned that you see coming in the future?

We expect the launch of PCM (Phase Change Memory) in big scale, and that will provide us much bigger memory storage. This will change the way we design the algorithms because now everything can be online rather than have to handle data from disk. Of course, PCM has its own problems such as read durability, and therefore we have to reduce the *reads and the writes on the PCM.*

Coming back to the performance issue, the relational databases in the cloud are facing a lot of competition from NoSQL and big data approaches. Do you think

that relational databases can compete with those two when high throughput is important?

We have to be clear that these two types of systems have been designed to meet different requirements and even different budget constraints. So they serve different market segments and therefore, there is no real competition. Of course, you can always extend relational database systems to handle what we have seen in a Web 2.0 application, but by that it's likely we end up with two engines running in parallel with some data sharing between the two engines.

I hear that you love B-trees compared to, for example, R-trees. What's wrong with R-trees?

That is not quite true. I love R-trees as well, but the problem with R-trees is the speed to process. Multiple traverses will take up much longer time during the locking process, so it incurs in a higher locking overhead.

That's true, but KD-trees, do you like them any better?

A KD-tree is fine, but the way it partitions the space does not lead to a balanced B+ tree.

So what trees do you like for spatial-temporal type…?

I like the B+ tree because it's very efficient, dynamic, and it's self-adaptive to load and data distribution.

But it doesn't work for multidimensional search, does it?!

Once you can linearize the data properly, then we can use the B+ tree.

You mean like a Z-ordering or something?

Or the Z-curve or we make use of distance just like the way we do it in iDistance, where we try to argue that even in high-dimensional space, I can measure one object against a reference point. If I can do that, that means that I divide the space into the Voronoi cells efficiently and that makes it very efficient. Then I can measure the distance to the reference point. Once I use that reference point plus the distance, I can index the data points using the typical B+ tree without changing the structures of the B+ tree.

I see. Okay, great. In the last five years, you've turned your attention from just writing papers to having startups. How do you make a database startup successful?

I set up my first startup in 1999 when we tried to provide photo-sharing systems to the public. That system works like Flickr except that we use keywords rather than tagging as well as comments for users to provide keywords to describe the photos, for users to share photos with a friend, and so on. But somehow it did not take off. So, to some extent, I did a startup 11 years ago, not five years ago.

Okay, okay, and more recently?

More recently, I had a startup on peer-to-peer systems based on my work on Bestpeer[1], and recently it has started to draw attention from venture capitalists, so I started to draw some business, but that is in China.

So do startups in China need to be entirely Chinese to be successful?

Not necessarily. But it's much harder to get the license to operate the business.

> ### *[...] for any department to be good in research, we need a huge number of good graduate students*

So, what can you say about the startup culture in China?

I think it's good because Chinese always like money. Just like me. I love money.

Okay, and is your startup in the Beijing area or ...?

In Hangzhou.

Are there many startups there now?

Hangzhou is a good place for startups. In fact, Alibaba has its headquarter and campus in Hangzhou, and one of the biggest web and game companies called NetEase is also housed in Hangzhou.

---

[1] http://www.bestpeer.com/

Okay, so I see now that you haven't gone back to the question of how do we make a database startup successful. So, is it like any other startup? Or different?

It's much harder to have a startup on trying to sell database systems because the market seems to be cornered by a few big players and in reality, it is very hard to do database migrations. So nobody is going to buy a new database system just because the new system has a few more functionalities or because it is cheaper since the data migration process is very expensive and time-consuming.

Does that mean your company aiming at customers that don't already have an existing DBMS?

The company that we have (Bestpeer, a P2P company) aims at customers who want to integrate details from different sources of databases and of course you can use Bestpeer like a middleware to link up different systems to share the data.

OK, great!

So, some of our readers told me: "Ten years ago we could seldom find papers from National University of Singapore (NUS) in tier 1 venues of the database community. Now, Beng Chin's database group is such an important part of the community pushing the frontiers of everyone else's research." Another reader said that you "built up NUS into a strong database research university starting from very little. I think this is an exceptional accomplishment, he hired everyone in the database group there, he aggressively recruited good students, particularly from China. He established a culture of publishing in the top places. He also actively mentored Ph.D. graduates, and all this was done well before he was dean. He was just a faculty member taking initiative and making things happen. Over time, he got institutional recognition and support including eventually his current deanship." So, long story but first question: what inspired you to do all this?

First of all, I just want to prove myself. Prove to myself that I could be just as good as any other database researchers. So in order to prove myself on that front, I had to publish where the top people published. And therefore even when I was doing my Ph.D., I started to focus on publishing at the top places. It's a fact that for any department to be good in research, we need a huge number of good graduate students. Therefore, we have to go out to look for good students. We are not MIT or Stanford, which can

attract good students without doing extensive outreach to these students. So for us, we have to go around and look for good students. We've convinced our colleagues in China, in Eastern Europe, and a few other countries to recommend good students to us.

So, for our readers who want to build up their own research group or department, how are you able to change the culture of your environment?

That is easy because no one wants to be a loser. So, the first step is to recruit good people and once they are in, provide them the environment, the freedom to do the research that they want to do, and support them in what they want to do. Set the goals and tell them what they are expected to do and explain how they will be rewarded. Once all this is made transparent, people tend to accept. Transparency reduces conflicts and workload for the administrators.

So then what were the most difficult challenges that you had to overcome to do this? Did they maybe oppose you on a rewards angle? Or what was the hardest part?

All were equally easy. [laughs]

OK, so let's see. What kind of rewards did you offer people when you talk of reward?

In our systems, we have a performance-based increment as well as performance-based bonus. That forms the rewards for those people who have done well for the last year and for the last three years. That really pushes people to work hard because everyone can be motivated to do well. Also, people are motivated when they are recognized.

I see. Did you have to base that off of that Clarivate SCI index or were you able to use other indicators of research quality and quantity?

We look at more than just the papers or the SCI index. We look at the impact a person makes to the research he's working on, his community, and his contributions.

So in the long run, I see how you can measure that impact, for example, it will show up in the H-index and everything, but you said last one year and last three years. So how can you measure the impact over that kind of time?

For the last one year, we look at how they perform in terms of the way they teach, the way they do their research, what papers they have got and what kinds of awards they've gotten for the last one year. For the three years window periods, we look at a much longer period where they could have made slightly longer impact for certain years, but they may not have the papers. We also look at their whole career since they joined the school.

> ## Set the goals and tell them what they are expected to do and explain how they will be rewarded.

OK. Database research in Asia. How is it different from database research in the rest of the world?

There's not much difference in the way we do research, but for Asians, it tends to be more algorithmic than system-oriented. So they build fewer systems because system development takes much longer time to materialize than just to focus on new problems and come out with solutions and show that they work.

Is that okay or is it a bad thing? That there's less systems research?

If we do not build systems, eventually we are driven to locate problems, and at times we end up creating artificial problems that could be new, novel, but on the other hand not applicable to the real world.

So system building is a reality check you would say.

Yes, that's true. There are problems and subproblems that evolve from the system. As we develop the system, we'll definitely encounter problems that cannot be managed or cannot be handled by the system that we tried to build.

OK. So I hear that you have played a role in raising the standard of database research in China. What have you done in that area and how do you see your role now?

I did not do as much as what has been said, but I do know most researchers in China very well and have talked with them often. I often advise them on what topics to move into and what not to do, how to protect the faculty's time, so that the faculty members can focus more on research.

Afternote: Beng Chin is an adjunct Chang Jiang professor at Zheajiang University.

So what should people not move into now?

They should move into user-driven research and try to build systems and address problems from the systems that they tried to develop for certain applications or for any other things that they have in mind to support.

And how can we protect faculty's time? What advice do you have for that?

For example, in China, they tend to organize more conferences. In order to organize conferences, you need faculty members to spend time organizing them and that will drain away a lot of a faculty member's time. Therefore, it's best to avoid organizing small conferences, those that have no consequences to the research quality. For NUS, we do not organize that many conferences. The last one we organized was VLDB, and we have not organized any for a long time.

OK. Would you recommend that new Computer Science Ph.D. graduates in the US and Europe consider a job in Asia?

Why not? Especially in Singapore. It's a very lovely and livable city. That's why you are here. [laughing].

To some degree, that's true. Your Ph.D. is from Australia, which follows the British system and the British offer this three-year Ph.D. with no coursework requirements. I've seen the kids coming out of these really short Ph.D. programs, and their resumes and abilities are about the same as that of a third-year graduate student in the U.S. On the other hand, U.S. Universities are very concerned because their Ph.D. students in CS can take five or six years to finish even though when they do finish they're ready to do research on their own. So at NUS, the CS degree's Ph.D. program is in between the US and the British system. You have personally worked to introduce courses and a written qualifying exam.

My question for you is: What is the right balance between the British system, which seems to me to be too short to teach people to do research, and then the American one, which takes a long time?

The graduate courses are necessary to provide a good foundation required for research. And of course, the qualifying exam (QE) is required since we take in students from different countries, with different standards, so we need the QE to weed out students who are not qualified enough to do Ph.D. With more courses to do, the students tend to take longer to graduate because they have to spend one year just to pass the courses and the QE, and have to do some courses in the second and third year. On the whole, it can take about four and a half years on average to graduate, but if a student works hard enough and works fast enough, he or she can still graduate within three years. It's not a problem. The real problem is those students who get very comfortable to stay on campus – after a while, they just do not want to leave and look for a job. If they can get an RA-ship, they hang on and just lead a student's life.

Is that bad? I mean you get to use them when they're really good at what they know: how to do research.

Yeah, that is good for faculty members. It may not be so good for the university because as a university, we want the students to leave as soon as they graduate because the resources are limited and we can use the resources to take in more students and educate more students.

OK. I've been told you're good at mining the strength of grad students even when the students are not very good. So how can you do good research with students who aren't very good? What's the secret? All of us have some students who are not so good. So how do we get good research out of students who are not so good?

That requires some time to understand the student's' strengths, to know what they can do, and from there we just ask them to do what they are good at. Of course, as a supervisor, I tend to guide them for the first few problems by telling them the likely solutions to solve certain problems. And after a while, they do learn the tricks, and as a requirement, they have to read ten papers a week in their first year and second year. And they have to write reports on the papers they read to me every Monday, and that builds up their foundation.

How long are the reports?

Just a few lines. It could be a couple of sentences, or it could be two or three paragraphs; but I told the students if they write the summaries based on the abstracts, I may call them up to get them to explain the papers to me.

So are these papers from all across the field? Or are these papers in the student's specific area?

They can read any papers they like, and sometimes I do point them to some papers that they have to read.

And so how long do you make them do this?

For the first two years.

That's a new tip for our readers.

So you have said that you run your students like an army. Another person told me "Beng Chin is a hard master to his students, very tough with them, he will scold them and may even check their code. However, he also rewards them, drinks and feasts with them, and plays with them. Basketball." And I believe it because apparently, that's a new basketball injury [pointing to his injured finger]. So even at the cost of your personal health, you play with them. My question is, how do you strike a balance between these two roles, which are quite different?

When it comes to working, it's about work. And we have to be serious about it; so we are very clear about our objectives, why we are here, and what are our goals. So we just want to just achieve that. The students know their role, and the students know my role, and therefore we work quite well when it comes to work. Once we start playing, there's no differentiation between supervisor and student. That's how I injured my two fingers.

So can you give an example goal that you would give a student?

Now, since I develop systems, all my students have to take part in developing the systems that I develop, even if they are not related to their thesis. I see it as part of their training. So we meet regularly to discuss the systems that are we building and of course they have to deliver what they are responsible for.

Afternote: since the interview, Beng Chin and his group have released Apache SINGA (a distributed machine learning engine), completed in-memory big data system, epiC, and a new storage engine, UStore, that supports blockchains and collaborative analytics.

I see. So a goal might be to write this module that satisfies this spec. So what kind of goals do you have that are not related to system building, that are more things that would maybe show up on their resume? Do you also have goals in that area?

We do. For their thesis, I tend to assign them a topic, and they focus on that topic and look for problems to solve. Quite often I show them the problems, but most of the topics revolve around the systems we're trying to build.

> **[my students] have to read ten papers a week in their first year and second year, write reports about those papers and send me every Monday [...]**

OK, got it.

Someone told me "Beng Chin drinks Chinese wine like a fish. Ask him which one is his favorite and whether he turns Chinese wine into good papers." Is there a conversion function?

I learned to drink Chinese wine only 10+ years ago when I started to go to China to recruit students. Initially, I disliked the taste, but after a while, I got used to it. Honestly, I don't like the taste of any alcohol. I just like the after effect.

Afternote: Beng Chin has stopped drinking for the Nth time.

[laughing] How does Chinese wine differ from other wine?

Chinese wine is very strong, and it's strong in smell too. It's very strong in the content of alcohol. Usually, the alcohol content is about 53-60%.

Wow. So we won't talk anymore about that. We'll have to go downstairs to that wine store in the basement after we do the interview.

So I hear that you are a skilled painter. What do you paint?

I did watercolor when I was young. I like painting since young. So when I had free time and when I felt inspired, I started to paint (when I was young). But when I came to Singapore, I took up Chinese painting. I attended night classes for three years. And I did draw some Chinese paintings, and one of the friends who got my painting is Raghu.

OK, congratulations Raghu! What did he have to do to get one?

He came at the right time.

One of Beng Chin Ooi's paintings

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

I will say take their work seriously, work hard, and once they put their heart into it, then they will do well. Work hard today and let tomorrow take care of itself.

OK, very good. Among all your past research, do you have a favorite piece of work?

That could be my Ph.D. work when I built a GIS system on top of an existing database system. So what I did was to extend an existing database system by developing a cartridge to handle GIS operations on top of it. That was my first system.

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

I would say charity work because it's a good way to contribute back to society.

So it would be charity work within IT aspect or charity…

Just for the general public because there are many people out there who need our help and our time.

Excellent. So, what charities in Singapore – or anywhere – are you involved with?

Embarrassingly, none. Therefore, I feel quite bad.

All right! If you can change one thing about yourself as a computer science researcher, what would it be?

I should not have stopped programming. I should have continued with my programming, and after stopping for so many years, now I'm not good at it.

Well if you had continued programming, you would have had less time to do all this other stuff. The startup, the department building, all that …

That is true. Because of my other roles, that's why I gave up programming. I used to program until 10 or so years ago.

Do you miss it?

Yep.

Maybe when you retire?

When I retire, I will open a restaurant. I will open a bar, be a bar attendant…

Oh, a restaurant and bar. What kind of food will you offer? Because I did hear you are a good cook. So what kind of food will be at your restaurant?

Chinese food. Because I know how to cook many dishes and I tend to do reverse engineering when I go to a restaurant, trying to figure out what are the ingredients, how do they cook, and I try to experiment with the food that I ate.

OK. Great.

And why I'd like to be a bar attendant? Because I like to drink and therefore I can serve as I listen to stories. [laughing]

Well great, thank you for telling me stories today.