YAHOO! RESEARCH

8/21/2009

Prof. Joseph M. Hellerstein
Department of Computer Science,
University of California, Berkeley
Berkeley, CA

Dear Prof. Hellerstein:

I read through the draft of your new proposal, "Data-Intensive Computing: Scalable Social Data Analysis," in response to the CFP from NSF's cross-cutting program on "Data-intensive Computing." As Chief Scientist for Yahoo!'s Audience division (the portal side of the company), I am acutely aware of the importance of data analysis and the impact it has on Yahoo!'s overall strategy as well as on specific product directions. We are fundamentally a data-driven company, and have made huge investments in gathering data that can shed light on trends that shape our competitive landscape and on users' product needs and preferences. We have many people within Yahoo! seeking to interpret this data. Further, many Yahoo! products, such as del.iciou.us, Flickr and Groups, intrinsically generate huge amounts of shared data, and are aimed at collaborative activities (e.g., tagging) centered around the data. I am therefore excited to see you embark on this project, which is squarely focused on an area of great interest to us.

The central premise of your proposal, that scalable data analysis requires social interactions and analysis tools that embed social context, is intriguing. For Yahoo!, any meaningful analysis must be scalable to extremely large datasets, in order to even be usable. But you are bringing to the fore an entirely different aspect of scalability, that of enabling a large number of people to collectively draw insights from shared data. On one level, this increases the ability of Yahoo! analysts to understand the data we gather. But the more exciting aspect of your work is that it could enable Yahoo! users to interact in richer ways online, and potentially lead to new product families that support social data analysis, or to new features of existing products such as Flickr. Personally, I believe that online "communities of purpose" are a key and nascent part of the exploding world of online communities. Yahoo! Answers, with over 100 million users, is an example of such a community of purpose, where the purpose is to help others in the community by providing answers to their questions. Your work might well lead to online communities whose purpose is to help others in understanding datasets, or to be data analysis collectives. Indeed, Swivel (which I know you have been closely involved it), Google Fusion, and many other emerging web communities are beginning to define this space. The work you propose can provide a rigorous foundation for such online communities and show how to deepen the interactions they support in cognitively well-founded ways. The impact can truly be huge.

We hope to attract graduate students from your project to do summer internships where we can pursue collaborative research, and look forward to following this work and providing feedback on your research plans and priorities. Our ongoing and past interactions give us a great starting point for a meaningful and sustained dialog.

In conclusion, allow me to provide some background on myself. I am a Research Fellow and the Chief Scientist for the Audience Technology and CCDI divisions at Yahoo!. (Audience is the portal side of Yahoo!, including properties such as the Front Page, Mail, Answers, Finance, Flickr, etc.; the other

2821 Mission College Blvd. Santa Clara CA 95054 ramakris@yahoo-inc.com

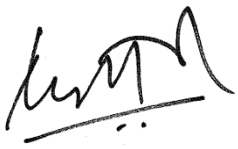
YAHOO! RESEARCH

8/21/2009

divisions are Advertising Platforms, CCDI, Search, and Mobile.) I'm on leave from University of Wisconsin-Madison, where I am Professor of Computer Sciences. I am Chair of ACM SIGMOD, on the Board of Directors of ACM SIGKDD and the Board of Trustees of the VLDB Endowment, and I have served as editor-in-chief of the Journal of Data Mining and Knowledge Discovery, associate editor of ACM Transactions on Database Systems, and the Database area editor of the Journal of Logic Programming. I am a Fellow of the Association for Computing Machinery (ACM), a Fellow of the Institution of Electrical and Electronics Engineers (IEEE), and I have received the Distinguished Alumnus Award from IIT Madras, a Packard Foundation Fellowship, an NSF Presidential Young Investigator Award, the ACM SIGKDD Innovations Award and the ACM SIGMOD Contributions Award.

Please contact me if I can be of help.

Sincerely,



Raghu Ramakrishnan
Research Fellow, Yahoo!
Chief Scientist for Audience and Cloud Computing

2821 Mission College Blvd. Santa Clara CA 95054 ramakris@yahoo-inc.com