Letter from the Editor-in-Chief

Evolution of the Bulletin

Over the years, the Bulletin has remained remarkably stable in its mission of informing readers about the state of ongoing work in important research areas– and making sure that best industry practices are made more widely known to the database research community. What has not remained completely stable is the infrastructure and what it enables in terms of delivering the contents of the Bulletin to readers.

The most recent example of this is that the program dvipdfm is now used to generate not only the full issue, but also individual papers of the issue. This resulted in my experiencing a "learning curve" (helped by Sudarshan) and required some modification of existing latex style files and tex files to accommodate this. The result is that now individual papers have correct page numbering. And this has been applied retro-actively back through 2010. My guess is that mostly authors will be sensitive to this, but that readers will notice it.

My expectation is that this kind of infrastructure evolution will continue, with resulting benifits for both authors and readers. If you have any suggestions about this, please let me know. The Bulletin is a continous work-in-progress. Having reader input in this would be extremely useful.

The Current Issue

We now live in what might be called a "post-SQL" world. SQL continues to be used widely. But many other avenues are being explored. This has been the result of living in a "post-DBMS" world. Again, not because DBMSs are no longer important. Quite the contrary. But what is new is that the domain of interest to query processing has expanded enormously- as the size of data subject to query processing has grown explosively.

This new world of dealing also with "big data" has seen the emergence of analytic engines, some based on map-reduce, some on streaming systems, etc. These systems typically are scan-oriented, processing enormous volumes of data at high speed. But if the data is large enough, even "high speed" may not produce results with an appropriately useful latency. Even were the software perfect, memory bandwidth provides serious limitations on system responsiveness.

The quest for useful data is the driver for many ML efforts. There is the intuition that if we can just find the data (assumed to be "out there", then we can solve interesting problems, both for industry and government.

Both these scenarios require a fresh look at data management. Enter "approximate query processing". This is not your father's Oldsmobile, or even your father's query processor. Though work has been going on that has nibbled at this space, this is an area that has bloomed in the current era. Thus, this is an excellent time to take a look at how work is proceeding in approximate query processing. This is, of course, a snapshot of the state-of-the-art. Work is ongoing in this area, with active, insightful, and important work continuing. Chris Jermaine has put together a very useful snapshot of the kinds of work that is being done in the research community. The issue is both an introduction and a survey of this area, with technology that is well worth knowing about, given the importance of the area. I want to thank Chris for serving as editor for the issue. It is well worth reading. This area will only grow in importance.

David Lomet
Microsoft Corporation