

Letter from the Editor-in-Chief

Technical Committee on Data Engineering

Kyu-Young Whang, our new Chair of the Technical Committee on Data Engineering has now appointed a new executive committee. Kyu-Young describes his appointments in his letter on page two. I have changed the inside front cover of the Bulletin to reflect Kyu-Young's appointments. These appointments reflect both continuity and new initiatives. I would urge you to read Kyu-Young's letter and become familiar with the new Executive Committee.

The Current Issue

There is no question but that we live in the age of "big data". I believe this is the result not only of new business models (online advertizing, etc.) but also of economics. The continuing decline in prices for storage and compute cycles means that ever larger amounts of both can be utilized for economic gain, as business pursue profits in a widening sphere of activity, both traditional and new. But it is not only of commercial interest. Science is now exploiting data at scales undreamed of in the past, for dealing with everything from searching for Higgs bosons to conducting vast environmental studies.

While distributed database systems have for a long time provided scalability, "big data" stretches well beyond what these systems have historically been able to handle. The map-reduce framework was the first to emerge to deal with the current vast flood of data. Not surprisingly, however, users always want more and better.

Map-reduce systems are now being augmented in fundamental ways. We are learning new ways (and some old ways) to describe what we want to do with our big data. Hence the re-emergence of SQL. With SQL and other high level ways of describing the way we wish to process big data, one needs a compelling story that performance will not suffer when using these idioms. Once again, query optimization enters the picture in a very important way.

The current issue of the bulletin addresses how the handling of big data has evolved. Sudarshan, as issue editor, has collected a good representative sample of work going on in this area. This includes industrial as well as university based approaches. The scale of data represents a technical challenge not faced in the past. But also a great commercial and scientific opportunity. So you can expect to see work continue in this area for many years to come.

I want to thank Sudarshan for the fine work he has done in assembling the issue. The issue should serve as a partial snapshot of the state of "big data" processing. But more, it could serve as a gateway to understanding how the field is evolving. I think you will enjoy the issue and greatly appreciate the efforts both of authors and of Sudarshan, the issue editor.

David Lomet
Microsoft Corporation