Cloud computing has been around long enough that we are all aware of it. However, it has not lived so long that major issues, technical and nontechnical, all have neat solutions. Cloud computing is at a stage similar to relational databases in the early 1980’s. Some technology exists, but there are many opportunities for improved technology and for turning the cloud into a profitable business. And, there is intense competition as vendors scramble to develop and deploy data centers and sell cloud services.

The unique technical characteristic of the cloud is that it provides compute power, data storage, and communication bandwidth on a scale that we have not really witnessed before. We now have enormous data centers supporting thousands of machines, attached by high speed, low latency communications, each with a number of attached high capacity disks.

The emergence of cloud computing is driven by economics. Cloud data centers are typically located where power is cheap and land costs low. Hardware is purchased in high volume at rock bottom prices or specially assembled from even cheaper components. Many data centers approach lights-out automated operations to drive down costs. Customers pay for what they use instead of provisioning for their maximum load. In addition, cloud providers offer their customers excellent availability via data replication. The economics of the cloud are compelling, and will produce industry wide changes.

The database community has a big role to play in exploiting cloud computing. We are, after all, in the business of efficiently storing and querying data. This special section on cloud computing covers a cross section of challenges posed by cloud computing. Query processing with Map-Reduce represents a very significant technical challenge. This special section has three papers in this area. There are, however other issues. The section also has one paper on monitoring the state (health) of cloud-based applications and another on how to price cloud services.

This section begins with the three query processing papers. “Optimizing Multiway Joins in a Map-Reduce Environment,” by Foto N. Afrati and Jeffrey D. Ullman focuses on Map-Reduce systems and the optimization of multiway joins, paying particular attention to communication costs. “MAP-JOIN-REDUCE: Towards Scalable and Efficient Data Analysis on Large Clusters,” by Dawei Jiang, Anthony K.H. Tung, and Gang Chen describes a new filtering-join-aggregation strategy to improve scalability for data analysis when using Map-Reduce. Finally, “Heuristics-Based Query Processing for Large RDF Graphs Using Cloud Computing,” by Mohammad F. Husain, James McGlothlin, Mohammad M. Masud, Latifur R. Khan, and Bhavani Thuraisingham looks at the special problems presented by large RDF graphs, and how to effectively query them.

Query processing is not the only cloud challenge. One would like to watch over your application and be kept informed of how things are going. This is called “state monitoring” and is treated in “State Monitoring in Cloud Datacenters,” by Shicong Meng, Ling Liu, and Ting Wang. And, of course, the cloud is also a business, and needs to be run like one. How one prices cloud services determines whether a cloud provider grows and prospers or withers and dies. How to determine appropriate pricing is the subject of the last paper “Optimal Service Pricing for a Cloud Cache,” by Verena Kantere, Debabrata, Gregory Francois, Sofia Kyriakopoulou, and Anastasia Ailamaki.

This special section was a direct result of Beng Chin Ooi asking me to do it, and I want to thank him for the invitation. It has been a great opportunity to learn more about this increasingly important topic. I need also to thank the authors, not only of the accepted papers, but all of those who submitted their work. Not only would special sections not be possible without the enthusiasm of authors, but progress in our field has an essential dependence on this enthusiasm. Finally, thanks are due to the reviewers who worked so hard to evaluate and, in some cases, assist authors in the process of making their papers the strong contributions that they are. I believe readers will be well rewarded for reading the papers in this special section. To them I say “Bon appétit!”

David Lomet
Guest Editor

David Lomet is a principal researcher managing the Microsoft Research Database Group. Earlier, he worked at Digital, IBM Research, and the Wang Institute. He has a CS PhD degree from the University of Pennsylvania. He is the author of more than 100 papers (two SIGMOD “best papers”) and has 45 patents. He has served on program committees (SIGMOD, PODS, VLDB, ICDE, etc.), was ICDE 2000 PC cochair, VLDB 2006 PC Core Track chair, and is on the ICDE Steering Committee, the VLDB Board, is TCDE Chair and has been an editor for the ACM Transactions on Database Systems, VLDB Journal, and Journal of Distributed and Parallel Databases. He is the Data Engineering Bulletin Editor-in-Chief, for which he received the SIGMOD Contributions Award in 2010. He has received the IEE Golden Core, Outstanding, and Meritorious Service Awards and is a fellow of the IEEE, the ACM, and the AAAS.