

Towards the End-to-End Design for Big Data Management in the Cloud: Why, How, and When?

Divy Agrawal

Department of Computer Science
University of California at Santa Barbara
Santa Barbara, CA 93106
agrawal@cs.ucsb.edu

Abstract

With the wide-scale adoption of cloud computing and with the explosion in the number of distributed applications and end-user devices, we are witnessing insatiable desire to build bigger-and-bigger systems that can serve hundreds of millions of end-users, are highly automated, and can collect enormous amounts of data in short periods of time. Often newer systems are implemented by integrating existing sub-systems that are already in use. A consequence of such a massive-scale integration is that it is very difficult to have a complete understanding of the overall system design. In fact, recent examples indicate that the only way to debug and test newer modules is to put them in live deployments that sometimes can lead to disastrous outcomes. In this talk, I will use some of the recent events in the context of Big Data and Cloud Computing as a motivation to argue that we need better methodologies for end-to-end system design for big data management in the cloud. I will then explore some well-known abstractions from distributed computing and databases as a means towards such a design and conclude with a contemplative question whether we can achieve such a goal or shall we leave it all to an automated self-learning and self-corrective oracle.

Biography: Dr. Divyakant Agrawal is a Professor of Computer Science and the Director of Engineering Computing Infrastructure at the University of California at Santa Barbara. His research expertise is in the areas of database systems, distributed computing, data warehousing, and large-scale information systems. From January 2006 through December 2007, Dr. Agrawal served as VP of Data Solutions and Advertising Systems at the Internet Search Company ASK.com. Dr. Agrawal has also served as a Visiting Senior Research Scientist at the NEC Laboratories of America in Cupertino, CA from 1997 to 2009. During his professional career, Dr. Agrawal has served on numerous Program Committees of International Conferences, Symposia, and Workshops and served as an editor of the journal of Distributed and Parallel Databases (1993-2008), and the VLDB journal (2003-2008). He currently serves as the Editor-in-Chief of Distributed and Parallel Databases and is on the editorial boards of the ACM Transactions on Database Systems and IEEE Transactions of Knowledge and Data Engineering. He has recently been elected to the Board of Trustees of the VLDB

Endowment and elected to serve on the Executive Committee of ACM Special Interest Group SIGSPATIAL. Dr. Agrawal's research philosophy is to develop data management solutions that are theoretically sound and are relevant in practice. He has published more than 320 research manuscripts in prestigious forums (journals, conferences, symposia, and workshops) on wide range of topics related to data management and distributed systems and has advised more than 35 Doctoral students during his academic career. He received the 2011 Outstanding Graduate Mentor Award from the Academic Senate at UC Santa Barbara. Recently, Dr. Agrawal has been recognized as an Association of Computing Machinery (ACM) Distinguished Scientist in 2010 and was inducted as an ACM Fellow in 2012. He has also been inducted as a Fellow of IEEE in 2012. His current interests are in the area of scalable data management and data analysis in Cloud Computing environments, security and privacy of data in the cloud, and scalable analytics over social networks data and social media.