Cloud Computing for Bioinformatics and Medical Informatics

Dr. Yi Pan
Regents' Professor of Computer Science
Interim Associate Dean and Chair of Biology
Georgia State University
Atlanta, Georgia, USA
http://www.cs.gsu.edu/pan

Abstract:

Cloud computing has become a popular computing platform for many scientific and business applications with intensive data or computing requirements. Many massive data applications including bioinformatics and medical informatics applications should be the ideal applications on cloud platforms. However, with the current cloud programming models, complicated bioinformatics algorithms cannot be implemented easily and executed efficiently on the many cloud platforms. In this talk, I will give a review of different massively parallel computing platforms and compare various computing domains and programming models on these platforms, their limitations and potential solutions. In particular, I will point out the shortcomings and limitations of current cloud computing programming models for typical scientific algorithms, and propose possible solutions. Current MapReduce and Spark models and their variants have succeeded in data-parallel applications such as database operations and web searching; however, they are still not effective for applications with a lot of data dependency. We propose several approaches to solving this problem through extension of current programming models, automatic translation from sequential codes to cloud codes, simple API and framework built on current cloud models, detection of data and task parallelism, and their efficient scheduling. In particular, three code translators for parallel computing on clouds (M2M, J2M, and J2S) will be introduced. We will show that cloud computing is an ideal computing models for bioinformatics and medical informatics applications such as genome wide association study, metagenomics sequence assembly, and protein-to-protein interaction network analysis. Some preliminary theoretical and experimental results will also be reported in this talk.

Biography:

Yi Pan is a Regents' Professor of Computer Science, an Interim Associate Dean and Chair of Biology at Georgia State University, USA. He is also a visiting Changjiang Chair Professor at Central South University, China. Dr. Pan received his B.Eng. and M.Eng. degrees in computer engineering from Tsinghua University, China, in 1982 and 1984, respectively, and his Ph.D. degree in computer science from the University of Pittsburgh, USA, in 1991. His profile has been featured as a distinguished alumnus in both Tsinghua Alumni Newsletter and University of Pittsburgh CS Alumni Newsletter. Dr. Pan's research interests include parallel and cloud computing, wireless networks, and bioinformatics. Dr. Pan has published more than 180 journal papers with over 60 papers published in various IEEE journals. In addition, he has published over 150 papers in refereed conferences. He has also co-authored/co-edited 40 books. His work has been cited more than 7000 times. Dr. Pan has served as an editor-in-chief or editorial board member for 15 journals including 7 IEEE Transactions. He is the recipient of many awards including IEEE Transactions Best Paper Award, several conference Best Paper Awards, IBM Faculty Award, JSPS Senior Invitation Fellowship, IEEE BIBE Outstanding Achievement Award, NSF Research Opportunity Award, and AFOSR Summer Faculty Research Fellowship. He has organized many international conferences and delivered keynote speeches at over 50 international conferences around the world.