Abstract:

In July 2010, we obtained a new record on computing specific bits of Pi, the mathematical constant. The new record consisted of 256 bits of Pi around the two quadrillionth bit position. The computation was carried out by a MapReduce program called DistBbp. DistBbp ran on Apache Hadoop clusters and utilized idle cluster resources without monopolizing the entire cluster. One of the largest computations took 23 days of wall clock time and 503 years of CPU time on a 1000-node cluster. In this talk, we discuss our experience on performing such computation and the lessons we have learned. Apache Hadoop is an open source distributed computing software system. It is now the core system behind many companies including Yahoo!, Facebook, Twitter, etc. We will give a brief introduction about Hadoop in this talk.

About the Speaker:

Dr. Tsz-Wo Nicholas Sze received his B.Eng. and M.Phil. degrees in Computer Science from the Hong Kong University of Science and Technology in 1999 and 2001, respectively. He received his Ph.D. degree in Computer Science from the University of Maryland College Park in 2007. Then, he joined Yahoo! Cloud Computing team in Silicon Valley, California, USA. His research interests include computational number theory and distributed computing. He is also a member of the Project Management Committee of Apache Hadoop, an open source distributed computing software system.