

所级学术报告会

报告题目: PortHadoop-R: Support the Merging of HPC and Cloud

时 间: 8月18日(周五)上午10:00

地 点: 计算所四层报告厅

报告人: Xian-He Sun, Professor,

Illinois Institute of Technology



摘要:

High Performance Computing (HPC) is becoming data intensive. In the meantime, big data applications are requiring more and more computing power. The merging of HPC and big data analytics is inevitable. However, the conventional HPC ecosystem, represented by MPI and Parallel File Systems (PFS) environments, and the newly emerged Cloud/big data ecosystem, represented by MapReduce/Spark and Hadoop File Systems (HDFS) environments, are designed for different applications and with different design principles. They do not work together naturally. Even worse, by the CAP theory, any of the two ecosystems cannot be extended to have all the merits of the other. In other words, these two ecosystems will co-exist. The best we can have is a merged system which can provide the functionality and merits of both ecosystems. In this study, we provide the PortHadoop-R solution to support the merging of HPC and Cloud at the file level. PortHadoop-R allows data to be read directly from PFS to the memory of Hadoop nodes and integrates the data transfer with R data analysis and visualization. PortHadoop-R is carefully optimized to utilize the merits of PFS and MapReduce to achieve concurrent data transfer and latency hiding. PortHadoop-R is tested on NASA climate modeling applications. Experimental results show PortHadoop-R delivered a 15x speedup. Even without the 15x speedup of PortHadoop-R, the MapReduce environment is already significantly faster than MPI clusters on processing climate data. PortHadoop-R further demonstrates the potential of the merging of HPC and Cloud.

报告人简介:

Dr. **Xian-He Sun** is a University Distinguished Professor of Computer Science of the Department of Computer Science at the Illinois Institute of Technology (IIT). He is the director of the Scalable Computing Software laboratory at IIT and a guest faculty in the Mathematics and Computer Science Division at the Argonne National Laboratory. Before joining IIT, he worked at DoE Ames National Laboratory, at ICASE, NASA Langley Research Center, at Louisiana State University, Baton Rouge, and was an ASEE fellow at Navy Research Laboratories. Dr. Sun is an IEEE fellow and is known for his memory-bounded speedup model, also called Sun-Ni's Law, for scalable computing. His research interests include data-intensive high performance computing, memory and I/O systems, software system for big data applications, and performance evaluation and optimization. He has over 250 publications and 5 patents in these areas. He is a former IEEE CS distinguished speaker, a former vice chair of the IEEE Technical Committee on Scalable Computing, the past chair of the Computer Science Department at IIT, and is serving and served on the editorial board of leading professional journals in the field of parallel processing. More information about Dr. Sun can be found at his web site www.cs.iit.edu/~sun/.