

Biomedical Big Data Science on Tianhe Supercomputers

Dr. PENG Shaoliang

Professor,

National University of Defense Technology, Changsha, China

DATE : 30 September 2016, Friday *(including Q & A session)*

TIME / VENUE : 1630 – 1800 / SCM809

LANGUAGE : English

FACILITATOR : Dr. ZHU Hailong

Abstract

Extremely powerful computers are needed to help scientists to handle high performance computational biology and drug design problems. The world's largest genomics institute BGI currently generates 6 TB data each day. The European Bioinformatics Institute (EBI) in Hinxton currently stores 20 petabytes (1 petabyte is 1015 bytes) of data and back-ups about genes, proteins and small molecules. TianHe supercomputers can speed up computational biology and drug design processing. In 2013, 2014, and 2015, Tianhe-2 topped the TOP500 list of fastest supercomputers in the world. Many well-known bioinformatics and drug design softwares (BWA, DOCK, SOAP3-dp, SOAPdenovo, SOAPsnp etc.) are developed and running on TH-2.



The talk focuses on two main areas: 1. Drug Design: mD3DOCKxb is a largest high throughput molecular docking platform and finishes the docking of all the purchasable molecules (about 42 million) on earth within 24 hours. It has a parallel efficiency of over 70% using 192,000 CPU cores and 1,368,000 MIC cores. It gains the Gold Award of PAC 2015 (Parallel Application Challenge Competition) and is reported by CCTV 1, 《ScienceNet》, 《China Science and Technology News》, and 2015 Top 10 News of Hunan Province of China. 2. Genetic Engineering: The “Human Whole Genome Re-sequencing Analysis Software Pipeline” is firstly designed by applicant. The whole analyzing procedure takes 4 hours to finish the analysis of a 300 TB dataset of whole genome sequences from 2,000 human beings. The speedup is about 1200X.

Tianhe supercomputers can handle 3 kinds of computational biology and drug design problems: computation intensive, memory intensive, and communication intensive. Now Tianhe supercomputers are open online to all the scientists not only in China but also all over the world.

Keywords: computational biology; drug design; TianHe supercomputer; high performance computing

Speaker

Dr PENG Shaoliang is a professor in National University of Defense Technology (NUDT, Changsha, China) and an adjunct professor of BGI. He was a visiting scholar at CS Department of City University of Hong Kong (CityU) from 2007 to 2008 and at BGI Hong Kong from 2013 to 2014. His research interests are high performance computing, bioinformatics, virtual screening, and biology simulation. He has participated in many keystone projects in China such as TianHe supercomputers. He gains the Gold Award twice of PAC 2014 and 2015 (Parallel Application Challenge Competition) and is reported by CCTV 1, 《ScienceNet》, 《China Science and Technology News》, and 2015 Top 10 News of Hunan Province of China (1. Human Whole Genome Re-sequencing Analysis Software Pipeline, 2. mD3DOCKxb: largest high throughput molecular docking platform). He also gains the Finalist Awards of 2015 IEEE International Scalable Computing Challenge, SCALE 2015. He has published 3 books and over 50 papers in ACM/IEEE Transactions, Nature Communications, Cell AJHG, Genome Biology, Cancer Research, BMC Bioinformatics, IPDPS, and SCIENCE CHINA. The downloading times of software are over 50000. He is Executive Editor and Associate Editors of several international journals (Interdisciplinary Sciences: Computational Life Sciences, IJCSE, IJHPCN, and IJES). Moreover, he is the PI of several key projects including 973, 863 and National Natural Science Foundation of China (NSFC).

****Welcome****