Mapping the Universe;
Challenge for the Big Data science

A Talk by

Dr.Utain Sawangwit

National Astronomical Research Institute of Thailand

utane@narit.or.th

ABSTRACT

Observational Cosmology relies on big surveys and very often full-sky reconstructions of the physical universe.

The many tools utilized by cosmologists span a wide range of techniques from observing the earliest radiation after the Big Bang to the use of gravitational lensing effect and supernovae explosions. In this talk I will review the challenge that modern cosmology has overcome and what it is facing in the near future when even larger telescopes hence bigger surveys and datasets start to roll in during the coming decades.

Workshop on e-Science and High Performance Computing (eHPC2015)
23 July 2015 at Prince of Songkla University , HatYai Campus
in conjunction with JCSSE 2015

DOM: A big data analytics framework for mining Thai public opinions

A Talk by
Assoc.Prof.Dr. Tiranee Achalakul
King Mongkut's University of Technology Thonburi

This work is a part of the "From Research to Product" program at the department of computer engineering

ABSTRACT

We, human being, have never been more connected through the emergence of social networks. Social networks, in terms of both data and users, have been exponentially growing and connect our lives together in various dimensions. We can connect with people across the planet with a touch of a finger. In every second, hundred thousands of messages are shared through social media such as Facebook, Twitter, Foursquare, and etc. They are about our life, feeling, experience and opinion. This practically represents the 21st century of our civilization, "The era of social network".

Social media networks generated huge volumes of data. They have been use in various types of applications including public health, emergency coordination, news recommendation, and stock market prediction. The data from social media networks gathered under the catch-all term, "big data". However, as much as 90% of the data stored is "unstructured," meaning that it is spontaneously generated and not easily captured and classified. Big data is only valuable if it tells a story. The organizations that can use stories to make sense of big data are going to excel.

In this talk, I will present the developments of DOM (Data and Opinion Mining), a big data analytics engine that is capable of mining Thai public opinions regarding specific issues discussed on the social network sites, and its corresponding mobile solution for answering public opinions about events and locations. The engine takes in data from multiple well-known social network sources, and then processes them using MapReduce, a keyword-based sentiment analysis technique, and an influencer analysis algorithm to determine public opinions and sentiments of certain topics. The system was evaluated its sentiment prediction accuracy by matching the predicted result with the human sentiment and tested on various case studies. The effectiveness of the approach demonstrates the practical applications of the engine.

Workshop on e-Science and High Performance Computing (eHPC2015)
23 July 2015 at Prince of Songkla University , HatYai Campus
in conjunction with JCSSE 2015

Data Management for the CMS Collaboration

A Talk by

Dr. Kittikul Kovitanggoon

Department of Physic, Faculty of Science, Chulalongkorn University

kovitang.cern@gmail.com

ABSTRACT

The Compact Muon Solenoid (CMS) at CERN has been recognized not only by its complexities of the detector constructions, operations and physics discoveries, but also the large quantity of data and computing resources. One of the main technical challenges of the experiment is to organize the enormous data volume collected and to efficiently process the data for physics explorations. It would be difficult to meet the computing and storage requirements at any one place. Since the most of CMS collaborators are around the world and have their own significant local resources, the CMS computing system and environment are built as a distributed system of computing services and resources that interact with each other. Tier-0 computing centre at CERN will be directly connected to the experiment for initial processing and data archiving, but considerable processing will be done at remote Tier-1 centres, and analysis will be done at Tier-2 and Tier-3 centres, all interconnected using Grid technology.

In this talk, I will present you about the data management for the CMS collaboration. Starting with data recording after the proton-proton collisions at the CERN's Tier-0, the presentation will explain you how the collaboration handles the large amount of data and distributes the data across the world to Tier-1, Tier-2, and Tier-3 via the Grid system. Chulalongkorn University (CU) as a member of the CMS experiment collaborating with National Electronics and Computer Technology Center (NECTEC) for the derverlopment of the CMS Tier-2 at National Science and Technology Development Agency (NSTDA) and Tier-3 at CU will also be presented.

Workshop on e-Science and High Performance Computing (eHPC2015)
23 July 2015 at Prince of Songkla University , HatYai Campus
in conjunction with JCSSE 2015

Reproducible e-Science Experiments with Docker

A Talk by

Asst.Prof.Dr. Chanwit Kaewkasi

Department of Computer Engineering, School of Engineering,

Suranaree University of Technology

Email: chanwit@sut.ac.th

ABSTRACT

e-Science involves studying complex computational research works. It is usually hard for reviewers to prepare the computing environment, which is exactly same as authors.

Nature Biotechnology announced that it seeks a new way to evaluate computational tools more consistent. The journal is exploring to adopt Docker, a software tool that allows authors to share their computing environment.

We discuss steps how to use Docker for e-Science research. We also discuss how our team at SUT Aiyara Cluster laboratory contribute to the Docker projects.

HPC Experiences from ASC15 Asian Student Supercomputer Challenge and Implication for HPC Education

A Talk by

Asst.Prof.Dr.Putchong Uthayopas, Assoc.Prof.Dr. Chantana Chantrapornchai
HPCNC, Department of Computer Engineering,
Faculty of Engineering, Kasetsart University
Email: putchong@ku.th, chantrapornchai@gmail.com

ABSTRACT

ASC Student Supercomputer Challenge is an annual supercomputing competition held mostly in China. The goal is to promoting HPC application and education worldwide. This year competition, ASC15, is the largest competition in the world with more than 150 universities registered. Our team from Kasetsart University has been chosen as one of the 16 worldwide teams to compete in the final round. Each team has to design, construct a small HPC system, and tune it for selected HPC application under power usage of 3000 Watts. This presentation will share our experiences in preparation, and competition which focus on how to maximize the performance of leading-edge HPC applications such as HPL, NAMD, WRF-Chem, and Palobos, Gridding, and HPCG while maintaining the power consumption under 3000 Watt. This includes the optimal HPC system design that balanced between extreme performance and power consumption and the HPC applications performance tuning. This experience is very valuable for educating future generation engineer in HPC areas.

Workshop on e-Science and High Performance Computing (eHPC2015)
23 July 2015 at Prince of Songkla University , HatYai Campus
in conjunction with JCSSE 2015

The Image Post-processing of Weather Forecast Models with High Performance Computing

A Talk by

Dr. Prattana Deeprasertkul

Hydro and Agro Informatics Institute (Public Organization)

prattana@haii.or.th

ABSTRACT

The Weather Research and Forecasting (WRF) modeling system, the Regional Atmospheric Modeling System (RAMS), The Regional Ocean Modeling System (ROMS), the Coupled WRF-ROMS modeling system, and Simulating WAves Nearshore **(**SWAN**)** modeling are designed and developed for understanding the climate system and forecasting the weather. These global models produce the various timescales and domains of weather forecasts outputs, every day. Finally, a large volume of images are generated to display the hourly and daily weather forecasting with the image post-processing of modeling systems. This image processing computation needs the capability of more large-scale computing. Therefore, the High Performance Computing architectures are required to process these large volumes of images. Moreover, a parallel processing implementation of the geometric image transformation algorithm will be applied to operate the models post-processing on a cluster computer.

Workshop on e-Science and High Performance Computing (eHPC2015)
23 July 2015 at Prince of Songkla University , HatYai Campus
in conjunction with JCSSE 2015