2015 年度ハイパフォーマンスコンピューティング授業概要

2015 High Performance Computing Class Overview

松岡　聡　　　Satoshi Matsuoka

2015 年 10 月 5 日（月）　　Monday, Oct 10, 2015

本年度は、ハイパフォーマンスコンピューティング(HPC)、並列処理, クラスタ計算、クラウド、の分野において、エクサスケールに向かうスパコンや大規模 IDC にとって非常に重要な「ビッグデータと HPC の融合」の学術的な成果の近年の論文をサーベイする。

This year's class will deal with "convergence of HPC and Big Data"

- 基本的に、各自がハイパフォーマンスコンピューティング(HPC)、並列処理, クラスタ計算、IDC、ビッグデータ、クラウド、の分野において、「ビッグデータと HPC の融合に関する学術的な成果」にまつわる特定のテーマの<u>英語の論文を 3 編</u>選ぶ。それらを発表の 2 週間前に、下記のメイリングリストに出典とアブストラクトを mail する。すると、松岡はそのうち一本を選択するので、それを自分の回に発表する。
A student is designated per class to present a paper regarding "convergence of Big Data and HPC". The designated student will pick 3 papers (written in English), and email their abstracts 2 weeks prior to the assigned class. We will pick one to present out of the 3, and the specified paper will be presented in detail in front of the class.

- ビッグデータと HPC の融合が近年様々な形で模索されている。これは HPC にまつわるデータ処理をビッグデータ技術を用いて大規模化・広帯域化・高速化するだけでなく、ビッグデータに固有な処理を HPC 技術を用いて高速化・大規模化する課題もある。
Convergence of HPC and Big Data are being explored in various ways. This includes acceleration of lage-scale data in HPC, but also utilizing HPC technologies to accelerate big data processing such as graph processing and deep learning.
  - テーマの例 Example Themes:
    - Architectures for accelerating the handling of extremely large-scale data in HPC, such as burst buffers
    - Frameworks for Pre- and Post- processing of large-scale simulations
    - Visualization of extremely large-scale data
    - Utilization of new non-volatile storage and memory
    - Acceleration of Large-scale Graph Algorithms on Supercomputing Platforms
    - Accelerating Deep Learning using GPUs and other HPC accelerators
    - Large-scale parallel filesystems and their algorithms
    - Utilization of Many-core processors such as GPUs and Xeon Phi for big data
    - HPC-class networks for Big Data
    - Applications of converged Big Data and HPC platforms

  - We will not deal with small-scale "big data", pure HPC, or purely theoretical big data issues such as advanced theory of deep learning.

- The students should search for the appropriate papers in conferences and journals that deal with HPC and big data such as ACM/IEEE Supercomputing, ACM ICS, ISC (International Supercomputing Conference), IEEE Big Data, ACM HPDC, IEEE IPDPS (www.ipdps.org), ICS, Cluster Computing, CCGrid(www.ccgrid.org) USENIX, IEEE Transactions on Big Data. They are easily searcheable using Google Scholar and ACM/IEEE digital library portals.
- Such conferences often have associated workshops that specifically deal with convergence issues, so search widely.
- Papers that are too commercial or deal with non-innovative technologies, just stating the state-of-practice, should be avoided.

- The paper must have been published in the last 6 years, e.g., including 2009 and after. Also, papers that have been selected prior by other students cannot be selected. This implies that you should present as early as possible. You cannot present your own paper or a paper from your group.
- The selected abstracts and papers should be discussed and announced in the mailing list as indicated below.
- Although one paper is to be presented in a principal fashion, all three selected papers should be read and introduced as related topics. References to all papers should be indicated.
- Each class will involve presentation by one speaker; there will be many questions that will be asked by the instructor, the TA, and the student audience. Questions will be asked to audiences as well from the instructor and the TA, so full attention should be paid during the class.
- The grades will be marked in the following fashion
  - Presentation: 50%
  - Report: 50%
  - For each class, if a student voluntarily asks a question or states a technical opinion, additional 5% credit will be added
- Presentation will be done online using Powerpoint. A PDF version should be prepared and sent to the mailing list hpc15@matsulab.is.titech.ac.jp A DAY PRIOR TO THE CLASS (Sunday) so that the student audience can refer to it during the presentation. The presentation will also be archived at http://matsu-www.is.titech.ac.jp/lecture/lecture-wiki/index.php?hpc2015
- The report should summarize the general topic covering and including ALL THREE PAPERS regarding the state of the art in HPC and Big Data convergence, and should be 10 pages in IEEE conference paper format.