

Parallel Processing

TNMT16_HK2_2018_2019

Ho Chi Minh City University of Technology

References

1. *Parallel Computing – theory and practice*, Michael J. Quinn, McGRAW-HILL, 1994.
2. *Parallel Programming: Techniques and Applications Using Networked Workstations and Parallel Computers*, Barry Wilkinson and Michael Allen, Second Edition, Prentice Hall, 2005.
3. *Distributed Algorithms*, Nancy Lynch, Morgan Kaufmann, 1997.
4. *Scalable Parallel Computing: Technology, Architecture, Programming*, Kai Hwang & Zhiwei Xu, McGRAW-HILL, 1997.
5. *Introduction to Parallel Computing*: https://computing.llnl.gov/tutorials/parallel_comp/#Designing
6. *Open MP*:
7. *MPI*: <http://www.mcs.anl.gov/research/projects/mpi/tutorial/>
8. *Xeon Phi Programming*:
9. *GPU Programming*
10. *Hadoop*:
11. *Spark*:
12. *Parallel Computing - theory and practice*: http://www.cs.cmu.edu/afs/cs/academic/class/15210-f15/www/tapp.html#_preface

Lectures (10 weeks)

Week (Lecture 150min/week & Lab 150min/week)	Content	Outcomes	Teaching & learning activities	Evaluation means
1	Chapter 1. Introduction to Parallel Computing <ul style="list-style-type: none"> – Introduction – Definitions – Parallel methods 	L.O.1 L.O.4	Lecture, class discussion, reading textbooks [1][2]	Final exam, assignment.
Lab 1	Lab: Introduction & multithreading			
2	Chapter 2. Abstract machines: PRAM & BSP Multithreading OpenMP	L.O.1.1 L.O.3.2 L.O.4	Lecture, class discussion, reading textbooks [1][2] [6]	Final exam, lab, assignment.
Lab 2	Lab: OpenMP (1) Xeon Phi			
3	Chapter 3. MPI (Message Passing Interface) <ul style="list-style-type: none"> – Concept – MPI program structure – Point-to-point communication – Collective communication 	L.O.1 L.O.2.1 L.O.3.1	Lecture, class discussion, reading textbooks [7]	Final exam, lab, assignment.
Lab 3	Lab: OpenMP (2) & Xeon Phi			
4	Chapter 4. Parallel machine architectures: <ul style="list-style-type: none"> – Flynn classifications – Pipeline, Processor array, Multiprocessor, Data flow computer 	L.O.1.1	Lecture, class discussion, reading textbooks [1][2]	Final exam, lab, assignment.

Week (Lecture 150min/week & Lab 150min/week)	Content	Outcomes	Teaching & learning activities	Evaluation means
	– Processor organizations			
Lab 4	Lab: MPI – Point-to-point communication			
5	Chapter 5. Speedup: – Amdahl – Gustafson	L.O.2	Lecture, class discussion, reading textbooks [1]	Final exam, lab, assignment.
Lab 5	Lab: MPI – Collective communication			
6	Chapter 6. Map/Reduce – Map/Reduce – Hadoop	L.O.1.2 L.O.1.3 L.O.2 L.O.3 L.O.4	Lecture, class discussion, reading textbooks [10]	Final exam, lab, assignment.
Lab 6	Lab: Hadoop 1			
7	Chapter 7. Parallel & distributed computing techniques (1) – EPC – Partition, Divide & Conquer – Pipeline	L.O.1 L.O.2.1 L.O.3	Lecture, class discussion, reading textbooks [2]	Final exam, lab, assignment.
Lab 7	Lab: Hadoop & Spark			
8	Chapter 8. Parallel & distributed computing techniques (2) – Synchronous computations – Asynchronous computations – Load balancing	L.O.1 L.O.2.1 L.O.3	Lecture, class discussion, reading textbooks [2]	Final exam, lab, assignment.
Lab 8	Lab: GPU			
9	Chapter 9. Parallel algorithms	L.O.1 L.O.2	Lecture, class discussion, reading textbooks [2]	Final exam, lab, assignment.
Lab 9	Lab: Algorithms 1			
10	Chapter 10. Review	L.O.1 L.O.4	Lecture, class discussion, reading textbooks	Final exam, lab, assignment.
Lab 10	Lab: Algorithms 1			

HPC Lab

- SuperNode-XP 24 nodes x (2 CPUs x 12 cores, 2 Xeon Phi x 61 cores, 512/256/128 GB RAM, 1 TB HD/SSD), Infiniband 56 Gbps
- GPUs system: P100, P4

Evaluation

- Final exam: 40% HK

- Project + Class exercises & Project seminar (required): 30% HK
- Lab: 30%

Contact:

- PGS. TS. Thoại Nam: namthoai@hcmut.edu.vn
- TS. Nguyễn Quang Hùng: nqhung@hcmut.edu.vn

Class:

- Lecture: Monday, 112H6, T2-4 (KSTN: B1-315, T2-4)
- Lab: H6

Dataset

https://hadoopilluminated.com/hadoop_illuminated/Public_Bigdata_Sets.html

Đề Project môn TTSS

MT16_HK2_2018_2019

Quy định:

1. Mỗi nhóm tối đa 4 sinh viên (trùng với nhóm Project seminar).
2. Các nhóm đăng ký danh sách và đề tài cho Trưởng lớp, **hạn cuối 21/1/2019**:
 - Ghi rõ họ tên, MSSV, email của các thành viên trong nhóm
3. Nộp báo cáo 2-4 trang mô tả nội dung đề tài thực hiện, nguồn dữ liệu, dự kiến kết quả, **hạn cuối 15/4/2019**
4. Nộp báo cáo sơ bộ lên website Bkel/Sakai khoa CSE về tiến độ và kết quả đạt được, **hạn cuối 30/4/2019**.
5. Nộp báo cáo cuối kỳ lên website Bkel/Sakai khoa CSE, **hạn cuối 15/5/2019**, tất cả các nhóm phải nộp:
 - Báo cáo tối đa 8 trang A4
 - Mã nguồn.

Đề 1: Trục quan hoá các phương thức lập lịch static, dynamic cho các threads trong OpenMP.

Lý thuyết:

- Tìm hiểu về lập trình OpenMP
- Tìm hiểu về lập lịch static, dynamic
- Tìm hiểu công cụ trục quan hoá.

Hiện thực:

- Viết chương trình
- Trục quan hoá và phân tích kết quả.

Đề 2: Đánh giá hiệu năng của Spark trong các bài toán phân tích dữ liệu lớn.

Lý thuyết:

- Tìm hiểu về Spark & thư viện về phân tích dữ liệu lớn & học máy
- Triển khai thực tế trên hệ thống máy tính mạnh SuperNode-XP.

Hiện thực:

- Chạy các ứng dụng để đánh giá hiệu năng Spark trên SuperNode-XP
- Phân tích kết quả.

Đề 3: Đánh giá hiệu năng của các thư viện về Deep Learning như TensorFlow, Torch, Caffe... trên các loại GPU cards khác nhau.

Lý thuyết:

- Tìm hiểu về Deep learning và các thư viện TensorFlow, Torch, Caffe
- Triển khai thực tế trên hệ thống máy tính mạnh SuperNode-XP dùng các loại GPU như P100, GTX 1080Ti, P4, trên CPUs.

Hiện thực:

- Chạy các ứng dụng để đánh giá hiệu năng các thư viện TensorFlow, Torch, Caffe trên SuperNode-XP với các loại GPU khác nhau
- Phân tích kết quả.

Đề 4: Viết chương trình giải bài toán Association Rules dùng multi-thread trên Xeon Phi dùng cơ chế Offload và OpenMP

Lý thuyết:

- Tìm hiểu giải thuật cho Association Rules
- Tìm hiểu cách lập trình trên Xeon Phi dùng cơ chế Offload và OpenMP.

Hiện thực:

- Viết chương trình

- Vẽ biểu đồ đánh giá hiệu năng (speedup) cho từng trường hợp với số lượng threads khác nhau
- Thử nghiệm sử dụng 2 Xeon Phi cards.

Đề 5: Viết chương trình Association Rules dùng multi-thread trên Xeon Phi dùng OpenCL

Lý thuyết:

- Tìm hiểu giải thuật cho Association Rules
- Tìm hiểu cách lập trình trên Xeon Phi dùng OpenCL.

Hiện thực:

- Viết chương trình
- Vẽ biểu đồ đánh giá hiệu năng (speedup) cho từng trường hợp với số lượng threads khác nhau
- Thử nghiệm sử dụng 2 Xeon Phi cards.

Đề 6: Viết chương trình Association Rules dùng multi-thread trên Xeon Phi dùng Intel Cilk Plus và Intel TBB (Thread Building Blocks).

Lý thuyết:

- Tìm hiểu giải thuật cho Association Rules
- Tìm hiểu cách lập trình trên Xeon Phi dùng Intel Cilk Plus và Intel TBB (Thread Building Blocks).

Hiện thực:

- Viết chương trình
- Vẽ biểu đồ đánh giá hiệu năng (speedup) cho từng trường hợp với số lượng threads khác nhau
- Thử nghiệm sử dụng 2 Xeon Phi cards.

Đề 7: Viết chương trình nhân ma trận kích thước 1.000x1.000, 10.000x10.000 và 100.000x100.000 (có trao đổi hàng/cột giữa các bộ xử lý) trên hệ thống máy tính ảo có giao tiếp 1Gbps, 10 Gbps (Gigabit Ethernet), 40 Gbps (Infiniband).

Lý thuyết:

- Tìm hiểu và viết chương trình nhân ma trận dùng MPI có trao đổi hàng cột
- MPI One-Sided Communication

Hiện thực:

- Viết chương trình
- Vẽ biểu đồ đánh giá hiệu năng (speedup) cho 2 trường hợp dùng hệ thống máy tính vật lý và máy tính ảo với số lượng máy tính khác nhau
- So sánh trường hợp giao tiếp thông thường và cách sử dụng One-Sided Communication.

Bài 8: Viết chương trình so trùng ảnh dùng “The Skein Hash Function Family” trên Hadoop

Lý thuyết:

- Tìm hiểu Hadoop
- Tìm hiểu “The Skein Hash Function Family”
- Giải pháp loại các ảnh trùng nhau trong một tập ảnh lớn

Hiện thực:

- Viết chương trình
- Đánh giá hiệu suất trên hệ thống thực.

Đề 9: Viết chương trình Association Rules trên GPUs

Lý thuyết:

- Tìm hiểu về lập trình GPU (CUDA)
- Tìm hiểu về giải thuật cho Association Rules.

Hiện thực:

- Viết chương trình
- Đánh giá hiệu năng (speedup) với số lượng core khác nhau.

Đề 10: sinh viên có thể đề xuất bài toán để giải như K-means, SVM (Support Vector Machines), bài toán trên Graph...

Đề 11: Thử nghiệm các thư viện hỗ trợ DL như Pytorch. Nội dung: Khảo sát các loại bài toán: dùng CUDA với GPU, các chương trình DL sử dụng thư viện/không sử dụng thư viện. Đo hiệu suất ứng dụng đặc trưng trên các GPU cards như Tesla M2090, GTX 960, GTX 1080,...

Đề 12: Thử nghiệm các thư viện hỗ trợ DL như Tensorflow. Nội dung: Khảo sát các loại bài toán: dùng CUDA với GPU, các chương trình DL sử dụng thư viện/không sử dụng thư viện. Đo hiệu suất ứng dụng đặc trưng trên các GPU cards như Tesla M2090, GTX 960, GTX 1080,...

Đề 13: Giải pháp đóng gói ứng dụng Machine Learning trong môi trường Docker. Thử nghiệm với 1 ứng dụng cụ thể (khuyến khích làm từ 2 ứng dụng càng tốt).

Đề 14: Giải pháp đóng gói ứng dụng Deep Learning trong môi trường Docker. Thử nghiệm với 1 ứng dụng cụ thể (khuyến khích làm từ 2 ứng dụng càng tốt).

- Link: <https://github.com/NVIDIA/nvidia-docker>

- Link: https://www.hpcwire.com/solution_content/ibm/cross-industry/ensuring-cross-cloud-compatibility-for-gpu-workloads/

Đề 15: Sử dụng kỹ thuật DL để khuyến nghị mua-bán trên mã cổ phiếu dựa trên dữ liệu chứng khoán VN. Nội dung: Tìm hiểu về Tensor flow. Hiện thực giải thuật DL và đánh giá dựa trên dữ liệu chứng khoán VN.

Tham khảo

Video

High Performance Computing made easy, <http://www.hpc.uva.nl/>

Algorithms on Xeon Phi (XP)

1. Strassen algorithm: https://www.singularis-lab.com/docs/materials/07_Shapovalov_Strassen_CKA.pdf
2. Fast Smith-Waterman: <http://xsw.sdu-hpcl.org>
3. PCIT algorithm: <https://utexas.influent.utsystem.edu/en/publications/optimizing-the-pcit-algorithm-on-stampedes-xeon-and-xeon-phi-proc>
4. Breadth-first search: <http://www.dislab.org/docs/bfs-phi-paper-eng.pdf>
5. Graph coloring:
http://www.sandia.gov/~egboman/papers/Deveci_coloring_ipdps16.pdf
6. Pattern matching: <http://sbac.lip6.fr/2014/session%206/1-BitParallel.pdf>
7. Sort: <http://cass-mt.pnnl.gov/docs/ia3-2013/2-3.pdf>
8. OpenFoam:
<https://www.nersc.gov/assets/Uploads/IXPUGISC15OpenFOAMTCSV6.pdf>
9. Kalman Filter:
<https://facultystaff.richmond.edu/~ggilfoyl/research/keegan2014DNP.pdf>
10. SU2: http://stanford.edu/~economon/docs/SU2_IPCC_SciTech2015_final.pdf
11. Conjunction Gradient method:
https://en.wikipedia.org/wiki/Conjugate_gradient_method

Deep learning on XP

1. Deep Neural Networks for Financial Market Prediction:
https://www.researchgate.net/publication/281685181_Implementing_Deep_Neural_Networks_for_Financial_Market_Prediction_on_the_Intel_Xeon_Phi