

Parallel Processing & Distributed Systems

MT2011

Ho Chi Minh City University of Technology

<http://www.cse.hcmut.edu.vn/~nam>

References

1. *Scalable Parallel Computing: Technology, Architecture, Programming*, Kai Hwang & Zhiwei Xu, McGRAW-HILL, 1997.(*)
2. *Parallel Computing – theory and practice*, Michael J. Quinn, McGRAW-HILL, 1994.(*)
3. *Parallel Programming: Techniques and Applications Using Networked Workstations and Parallel Computers*, Barry Wilkinson and Michael Allen, Second Edition, Prentice Hall, 2005.
4. *Distributed Systems: Concepts and Design*, George Coulouris, Jean Dillimore, Tim Kindberg, Addison-Wesley, 2005.(*)
5. *Distributed Algorithms*, Nancy Lynch, Morgan Kaufmann, 1997.
6. *Distributed Operating Systems*, Andrew S. Tanenbaum, Prentice Hall, 1990.
7. *MPI*: <http://www.mpi-forum.org/docs/docs.html> (LAM)
8. *PVM*: http://www.csm.ornl.gov/pvm/pvm_home.html
9. *The GRID2: Blueprint for a New Computing Infrastructure*, Ian Foster and Carl Kesselman, Morgan Kaufmann 2004.

Lectures

Lectures (90 min)	Date	Topics	References	Notes
1	W1-1	Introduction	[1]	
Lab 1	W1-2	Lab: Multithread	[3]	
2	W2-1	Abstract machine models – PRAM, BSP, and parallel phase	[1][2]	
Lab 2	W2-2	Lab: MPI – Point-to-point communication		
3	W3-1	Parallel machine architectures: – Flynn classifications – Pipeline, Processor array, Multiprocessor, Data flow computer	[1] [2]	
Lab 3	W3-2	Lab: MPI – Collective communication 1	[3]	
		Pipeline	[1][2]	
4	W4-1	Speedup: – Amdahl – Gustafson – Sun and Ni's law	[1][2]	
Lab 4	W4-2	Lab: MPI – Collective communication 2	[3] [7]	
5	W5-1	Processor organizations, mapping and Scheduling	[1]	
Lab 5	W5-2	Lab: Master/slave model		
6	W6-1	Parallel paradigms and programming models	[2][3]	
Lab 6	W6-2	Lab: Speedup		
7	W7-1	Algorithms	[1]	
Lab 7	W7-2	Lab: Matrix algorithms		
Midterm: W8-9				
8	W8-1	Distributed systems: Introduction & Communication	[4]	
Lab 8	W8-2	Lab: Programming tools		
9	W9-1	Seminar: Name service	[4]	Ref
Lab 9	W9-2	Lab: Cluster	[3]	
10	W10-1	Seminar: Distributed File Systems	[4]	Ref
Lab 10	W10-2	Lab: VN-Grid	[3]	
11	W11-2	Seminar: Parallel & distributed computing techniques	[3]	Ref
Lab 11	W11-2	Lab: Parallel & distributed computing techniques	[3]	
12	W12-1	MapReduce & Hadoop	[3]	
Lab 12	W12-2	Lab: MapReduce & Hadoop		

13	W13-1	Seminar: Time and Global States	[4]	Ref
Lab 13	W13-2	Lab: Time and Global States	[3]	
		Web service	[4]	Ref
		Peer-to-peer systems	[4]	Ref
14	W14-2	Review		5
Lab 14	W14-2	Lab: Web service	[3]	
		Distributed algorithms (1)	[5]	Ref
		Distributed algorithms (2)	[5]	Ref

Evaluation (dự kiến)

- Midterm exam: 20%
- Final exam: 50%
- Exercises + seminar: 30% (project deadline week 13, review week 15)
- Contact email: hungnq2@cse.hcmut.edu.vn

ĐỀ BÀI TẬP môn XLSS&HPB

KTN2011

Quy định:

1. Mỗi nhóm tối đa 3 sinh viên (trùng với nhóm seminar).
2. Các nhóm đăng ký danh sách và đề tài cho Trưởng lớp trong vòng 2 tuần, **hạn cuối 4/9/2013**:
 - Ghi rõ họ tên, MSSV, email của các thành viên trong nhóm
3. Cuối học kỳ, **hạn cuối 02/12/2013**, tất cả các nhóm phải nộp:
 - Báo cáo tối đa 4 trang A4.
 - Mã nguồn

Bài 1: 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 thực nghiệm.

Đề 2: Viết chương trình nhân ma trận kích thước 1.000x1.000, 10.000x10.000 và 100.000x100.000 dùng multi-thread.

Lý thuyết:

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

Hiện thực:

- Viết chương trình
- Vẽ biểu đồ đánh giá hiệu năng (speedup) cho từng trường hợp với số lượng threads khác nhau.

Đề 3: 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, 40 Gbps.

Lý thuyết:

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

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.

Đề 4: MPI Deadlock

Lý thuyết:

- Tìm hiểu về Deadlock.

Hiện thực:

- Hiện thực giải pháp phát hiện Deadlock trong chương trình MPI.

Đề 5: Viết chương trình nhân ma trận kích thước lớn 1.000x1.000, 10.000x10.000 và 100.000x100.000 trên GPUs

Lý thuyết:

- Tìm hiểu về lập trình GPU (CUDA)
- Tìm hiểu về giải thuật nhân ma trận

Hiện thực:

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

Bài 6: Tạo vector timestamp cho các events `MPI_Send()`, `MPI_Recv()` và `MPI_Barrier()` trong chương trình MPI để tìm Causal distributed breakpoints.

Lý thuyết:

- Tìm hiểu Virtual clock

Hiện thực:

- Hiện thực việc tạo các vector timestamp cho các events như `MPI_Send()`, `MPI_Recv()` và `MPI_Barrier()`.
- Phân tích để tìm Causal distributed breakpoints.

Đề 7: Race condition

Lý thuyết:

- Tìm hiểu về race condition.

Hiện thực:

- Hiện thực giải pháp phát hiện race condition trong chương trình MPI.