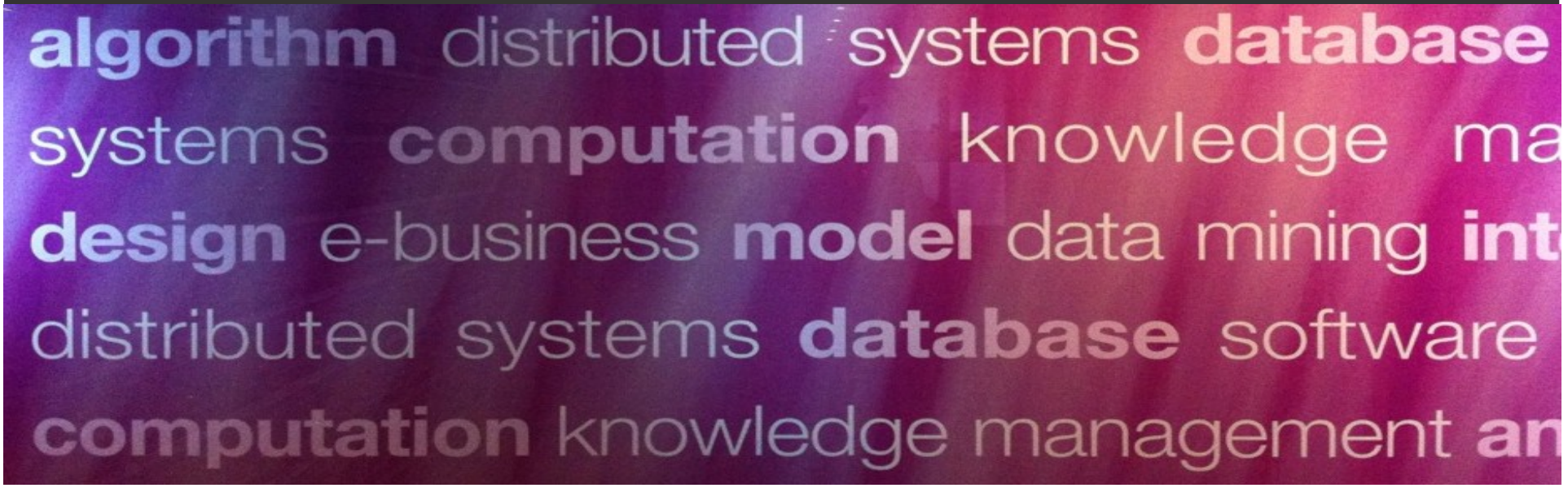
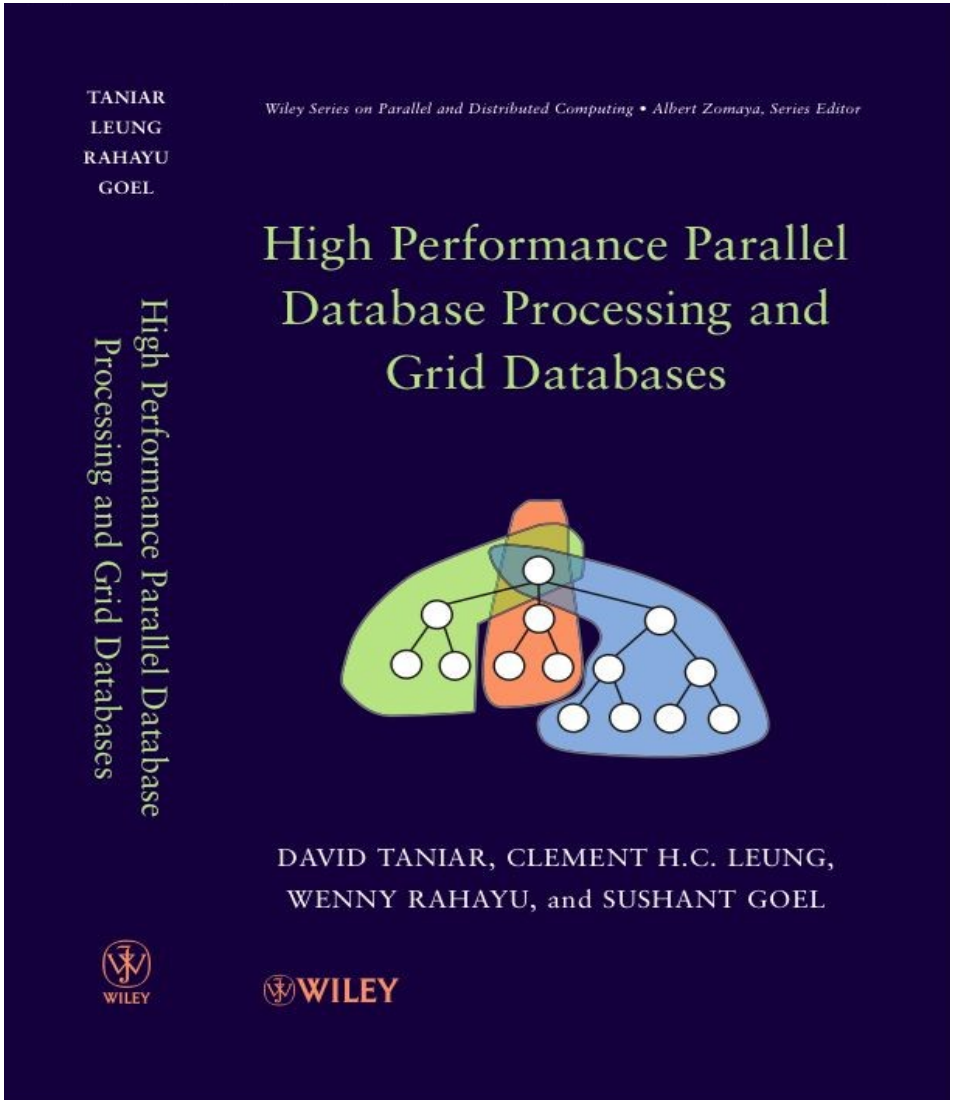
MONASH University

Information Technology

FIT5202 (Volume I)

Week 1c – Introduction to Parallel Databases





Chapter 1 Introduction

1.1/1.2. A Brief Overview, and Motivations

- An example:
  - If we have 1 petabyte of data, and the processing speed is 1GB/sec
  - How long does it take to process 1 PB of data?

1.1/1.2. A Brief Overview, and Motivations (cont'd)

- What is parallel processing, and why not just use a faster computer ?
  - Even fast computers have speed limitations
  - Limited by speed of light
  - Other hardware limitations

- Parallel processing divides a large task into smaller subtasks
- Database processing works well with parallelism (coarse-grained parallelism)
- Lesser complexity but need to work with a large volume of data

1.1/1.2. A Brief Overview, and Motivations (cont'd)

- **Moore's Law:** number of processors will double every 18-24 months
- CPU performance would increase by 50-60% per year
- Disk access time or disk throughput increases by 8-10% only
- Disk capacity also increases at a much higher rate
- I/O becomes a bottleneck
- Hence, motivates parallel database processing
- And parallel database processing is the foundation of **Big Data Processing**

1.3. Objectives

- The primary objective of parallel database processing is to gain performance improvement
- Two main measures:
  - **Throughput**: the number of tasks that can be completed within a given time interval
  - **Response time**: the amount of time it takes to complete a single task from the time it is submitted
- Metrics:
  - Speed up
  - Scale up

Exercise 1 (FLUX Quiz)

- Using the freeway analogy, number of cars that can pass through the freeway (M1: Monash Freeway) during the morning peak hour from 7 to 9am is called:

- A. Throughput
- B. Response Time
- C. None of the above
- D. A and B

Exercise 2 (FLUX Quiz)

- Using the freeway analogy, the duration I take to drive my car to go to work on a freeway (say M1 Monash Freeway) from the Burke Road entrance to the Blackburn Road exit is called:

- A. Throughput
- B. Response Time
- C. None of the above
- D. A and B

1.3. Objectives

- The primary objective of parallel database processing is to gain performance improvement
- Two main measures:
  - Throughput: the number of tasks that can be completed within a given time interval
  - Response time: the amount of time it takes to complete a single task from the time it is submitted
- Metrics:
  - **Speed up**
  - **Scale up**

1.3. Objectives (cont'd)

- **Speed up**
  - Performance improvement gained because of extra processing elements added
  - Running a given task in less time by increasing the degree of parallelism

- Linear speed up: performance improvement growing linearly with additional resources
- Superlinear speed up
- Sublinear speed up

$$\text{Speed up} = \frac{\text{elapsed time on uniprocessor}}{\text{elapsed time on multiprocessors}}$$

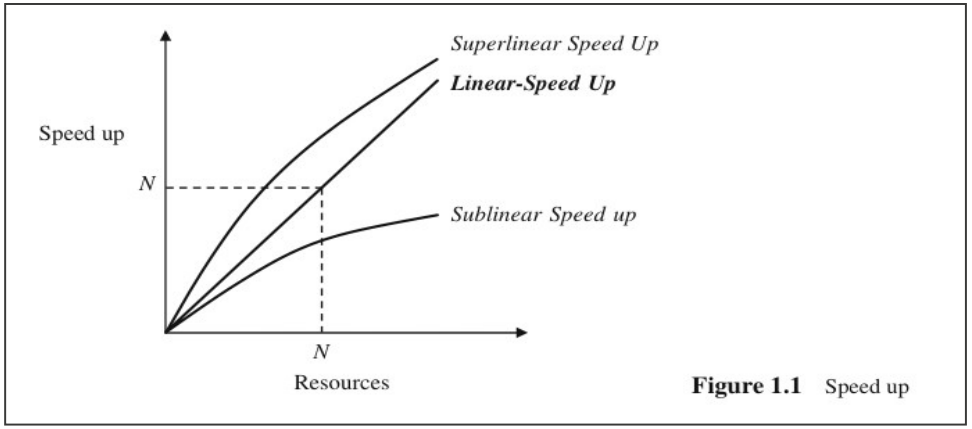


Figure 1.1 Speed up

1.3. Objectives (cont'd)

- **Scale up**
  - Handling of larger tasks by increasing the degree of parallelism
  - The ability to process larger tasks in the same amount of time by providing more resources.

- Linear scale up: the ability to maintain the same level of performance when both the workload and the resources are proportionally added
- Transactional scale up
- Data scale up

$$\text{Scale up} = \frac{\text{uniprocessor elapsed time on small system}}{\text{multiprocessor elapsed time on larger system}}$$

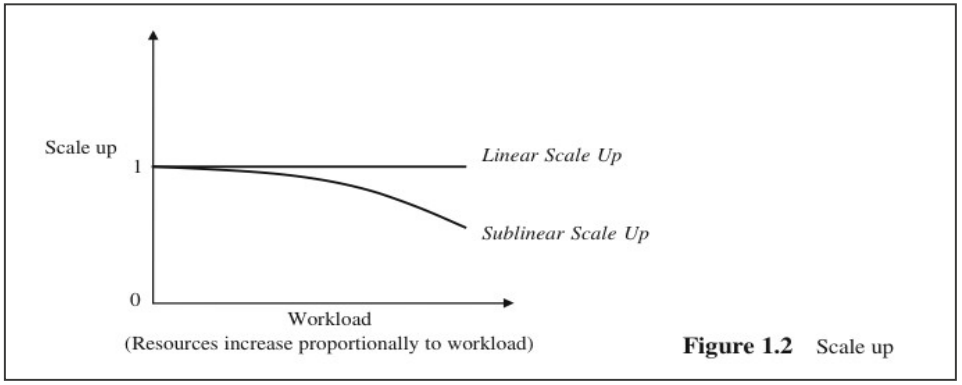


Figure 1.2 Scale up

1.3. Objectives (cont'd)

- **Transaction scale up**
  - The increase in the rate at which the transactions are processed
  - The size of the database may also increase proportionally to the transactions' arrival rate
  - *N*-times as many users are submitting *N*-times as many requests or transactions against an *N*-times larger database
  - Relevant to transaction processing systems where the transactions are small updates

- **Data scale up**
  - The increase in size of the database, and the task is a large job who runtime depends on the size of the database (e.g. sorting)
  - Typically found in online analytical processing (OLAP)

Exercise 3 (FLUX Quiz)

- **Using the current processing resources, we can finish processing 1TB (one terabyte) of data in 1 hour. Recently the volume of data has increased to 2TB and the management has decided to double up the processing resources. Using the new processing resources, we can finish processing the 2TB in 60 minutes.**

  Is this speed up or scale up? (5 Minutes)

- A. Scale Up
- B. Speed Up

1.8. Summary

- **Why**, **What**, and **How** of parallel query processing:
  - Why is parallelism necessary in database processing?
  - What can be achieved by parallelism in database processing?
  - How parallelism performed in database processing?
  - What facilities of parallel computing can be used?

Remember…

- There is only one question:
  - **Do you really want to pass this unit?**

Homework: Read Chapter 1 and Chapter 3 for next week