

IT4371: Distributed Systems

Overview and Introduction

|

Lecture slides by Dr. Nguyen Binh Minh

Department of Information Systems

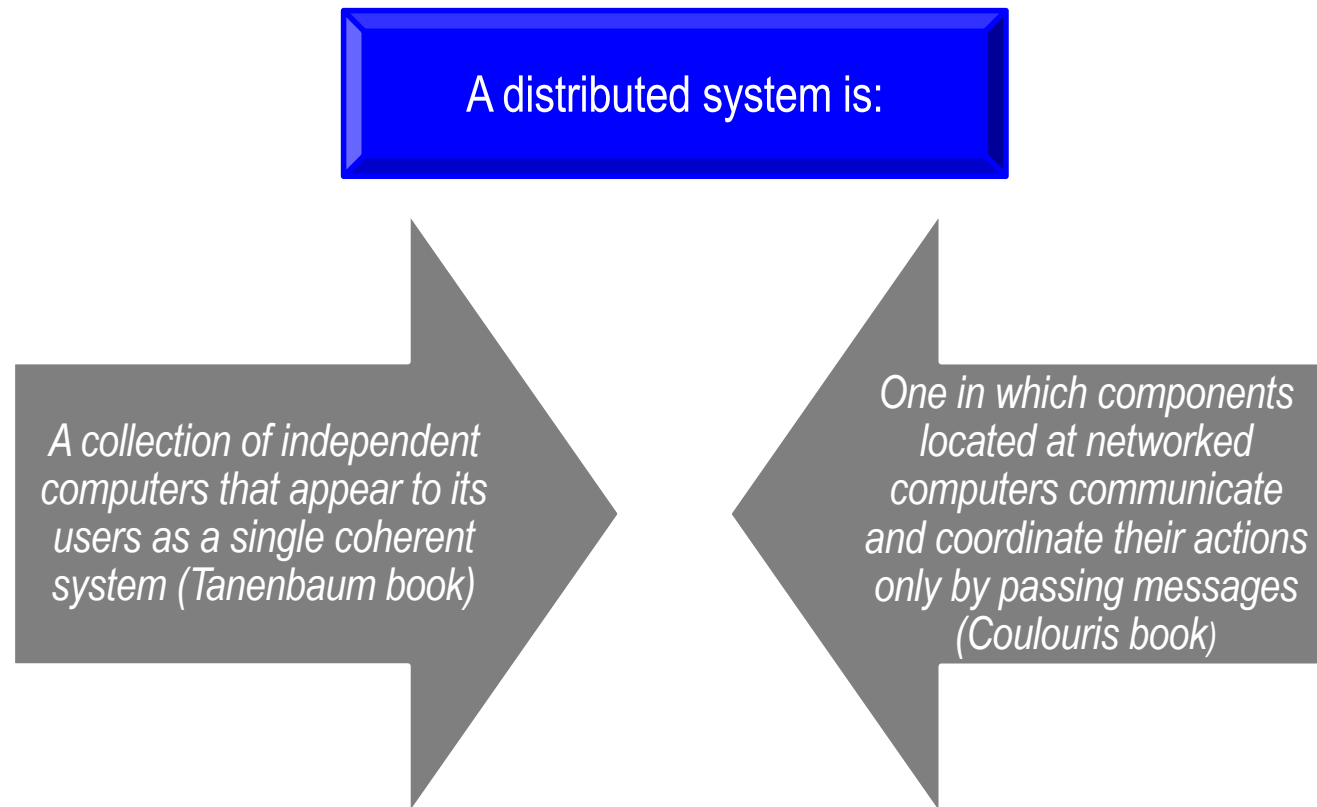
School of Information and Communication Technology

Hanoi University of Science and Technology

Why should you Study Distributed Systems?

Application Domain	Associated Networked Application
<i>Finance and commerce</i>	eCommerce e.g. Amazon and eBay, PayPal, online banking and trading
<i>The information society</i>	Web information and search engines, ebooks, Wikipedia; social networking: Facebook and MySpace.
<i>Creative industries and entertainment</i>	online gaming, music and film in the home, user-generated content, e.g. YouTube, Flickr
<i>Healthcare</i>	health informatics, on online patient records, monitoring patients
<i>Education</i>	e-learning, virtual learning environments; distance learning
<i>Transport and logistics</i>	GPS in route finding systems, map services: Google Maps, Google Earth
<i>Science</i>	The Grid as an enabling technology for collaboration between scientists
<i>Environmental management</i>	sensor technology to monitor earthquakes, floods or tsunamis

Definition of a Distributed System



Why Distributed Systems?

Scale

- Processing Data

Diversity in Application Domains

Collaboration

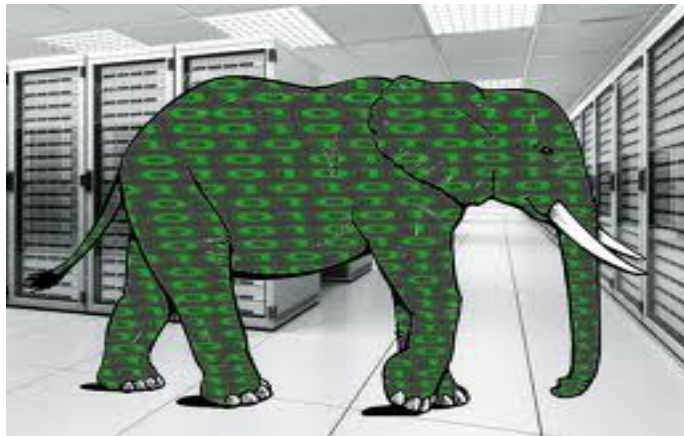
Cost

Why Distributed Systems?

A. *Big data* continues to grow:

- In mid-2010, the information universe carried 1.2 zettabytes and 2020 predictions expect nearly 44 times more at 35 zettabytes coming our way.

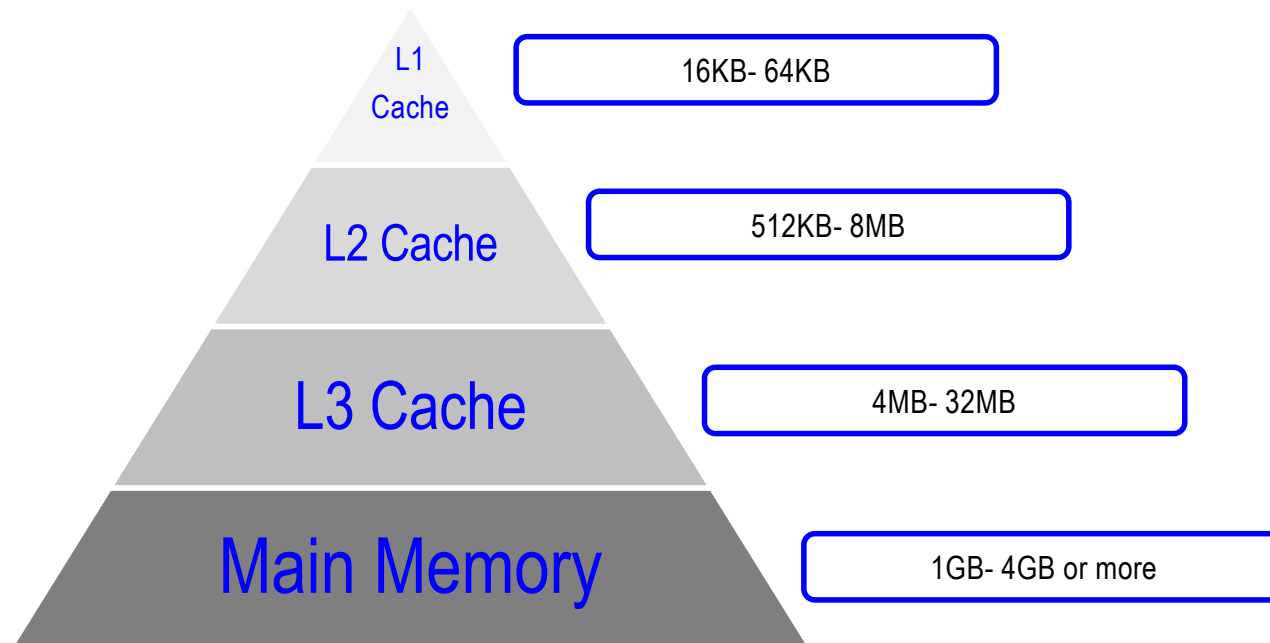
B. Applications are becoming *data-intensive*.



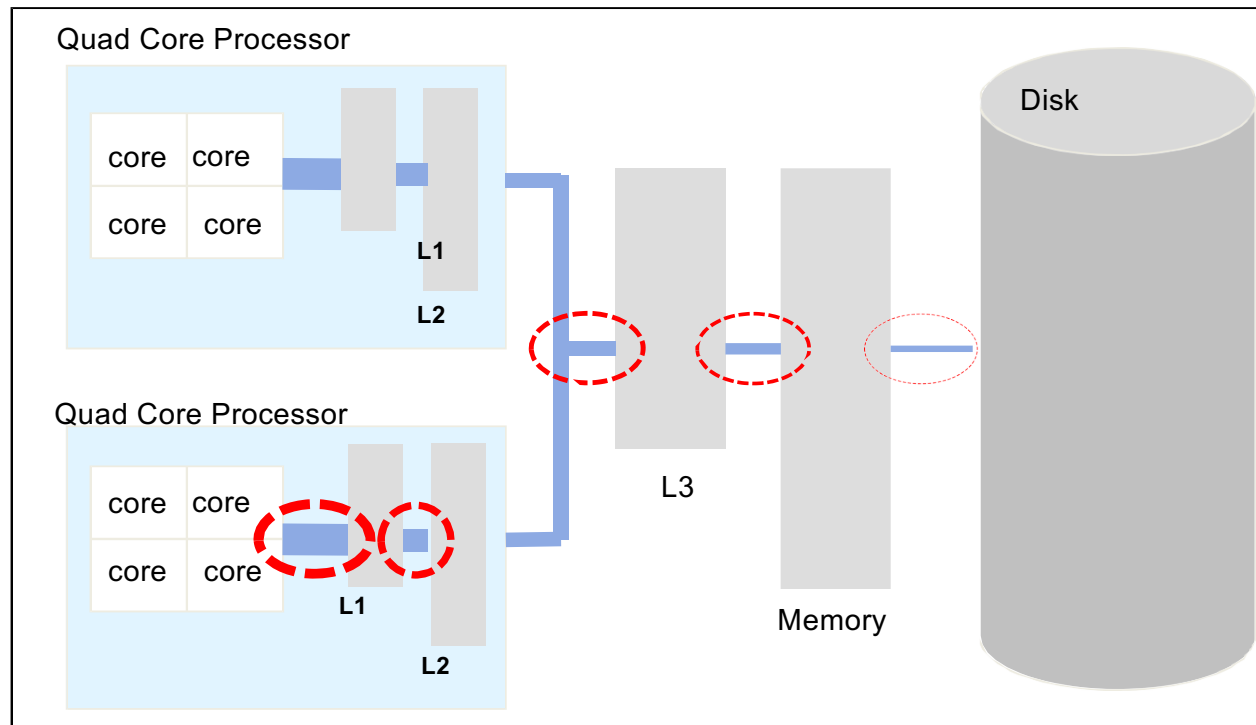
Why Distributed Systems?

C. Individual computers have limited resources compared to scale of current day problems & application domains:

1. Caches and Memory:



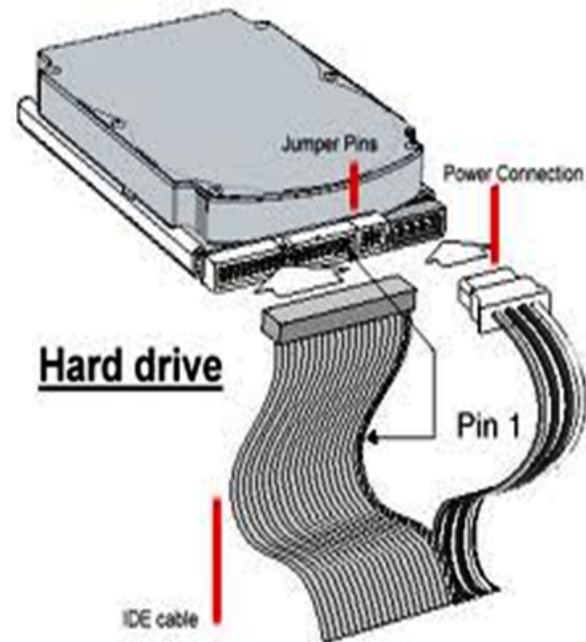
Blade Performance



Why Distributed Systems?

2. Hard Disk Drive:

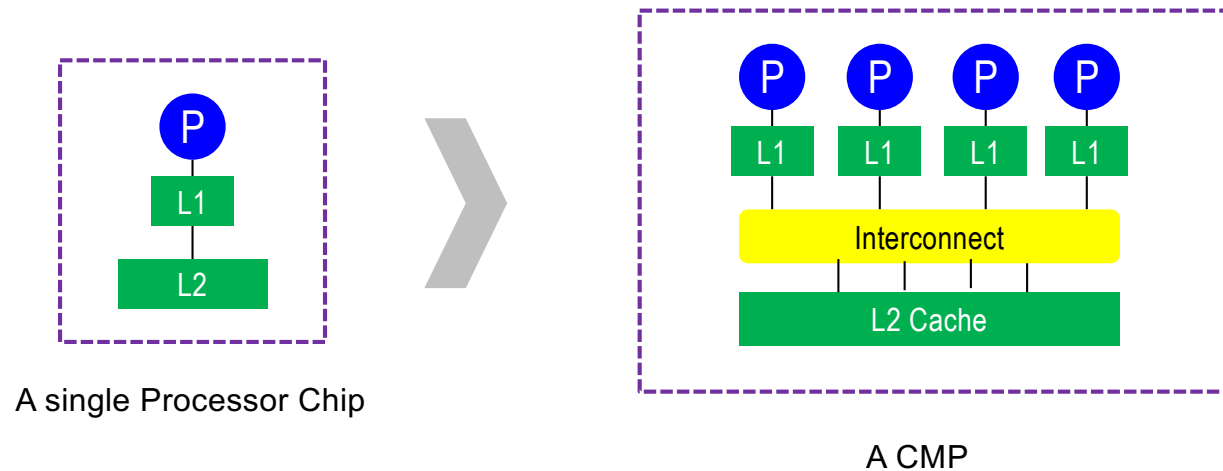
- Limited capacity
- Limited number of channels
- Limited bandwidth



Why Distributed Systems?

3. Processor:

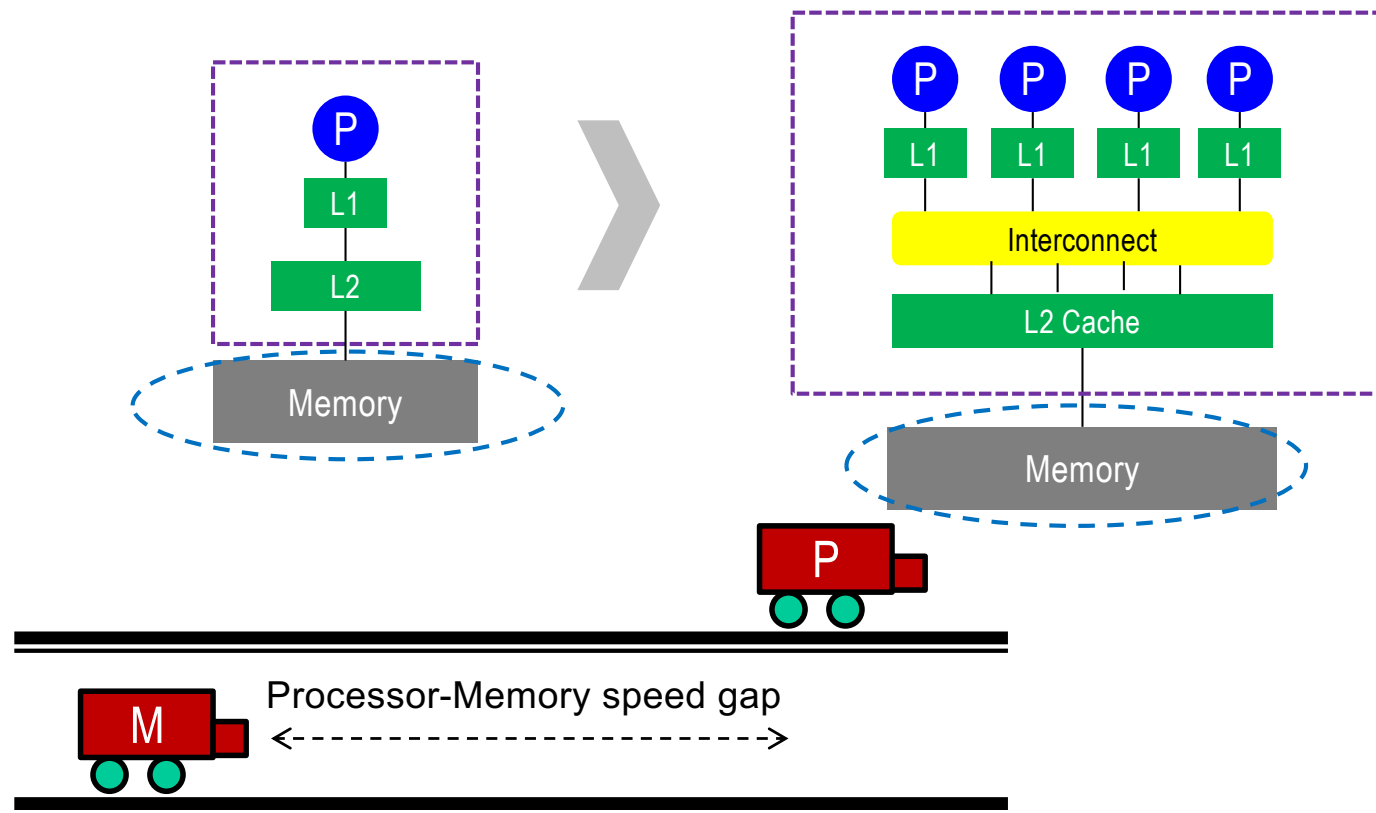
- The number of transistors that can be integrated on a single die has continued to grow at Moore's pace.
- Chip Multiprocessors (CMPs) are now available



Why Distributed Systems?

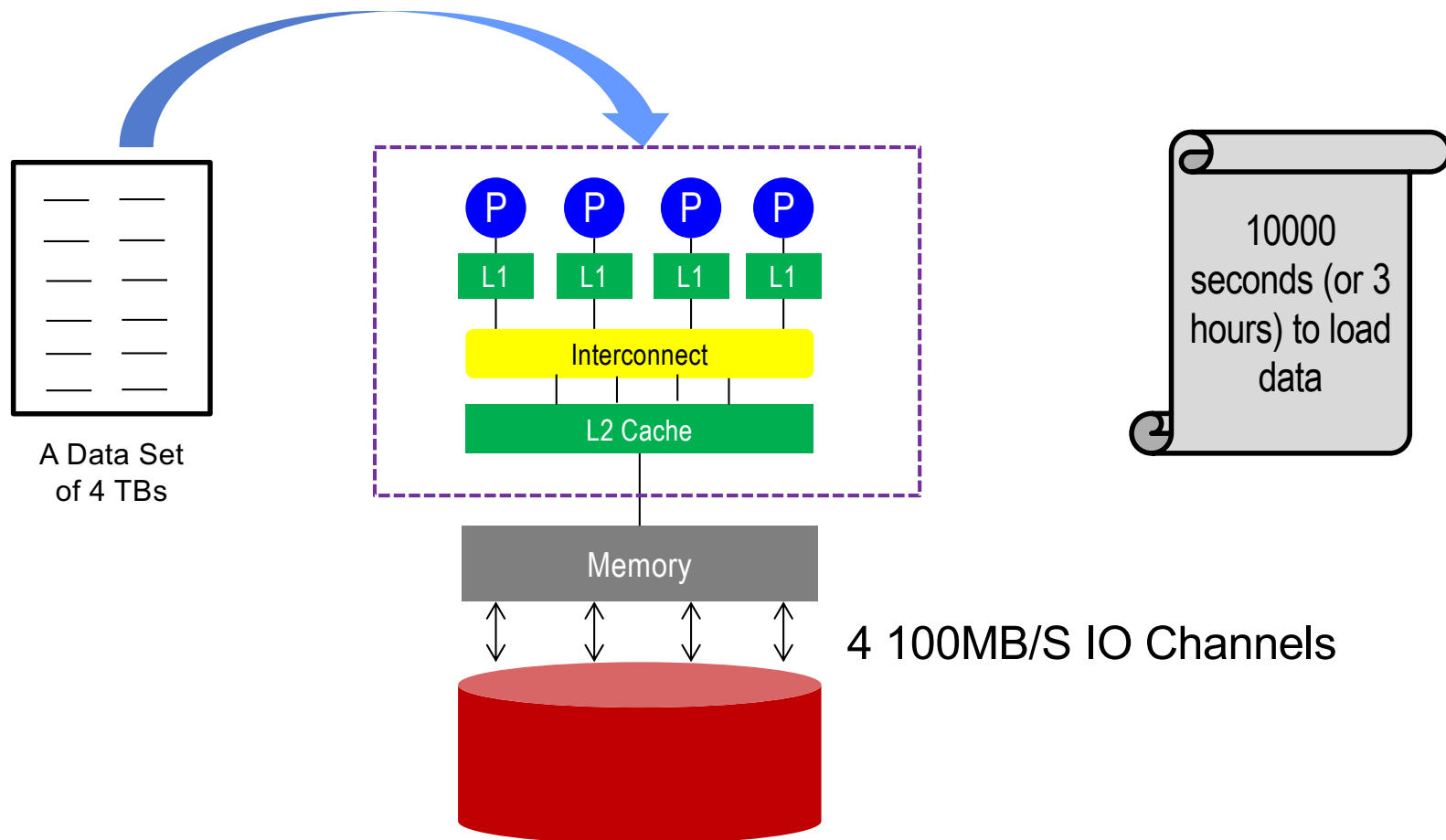
3. Processor (cont'd):

- Up until a few years ago, CPU speed grew at the rate of 55% annually, while the memory speed grew at the rate of only 7% [HP report].

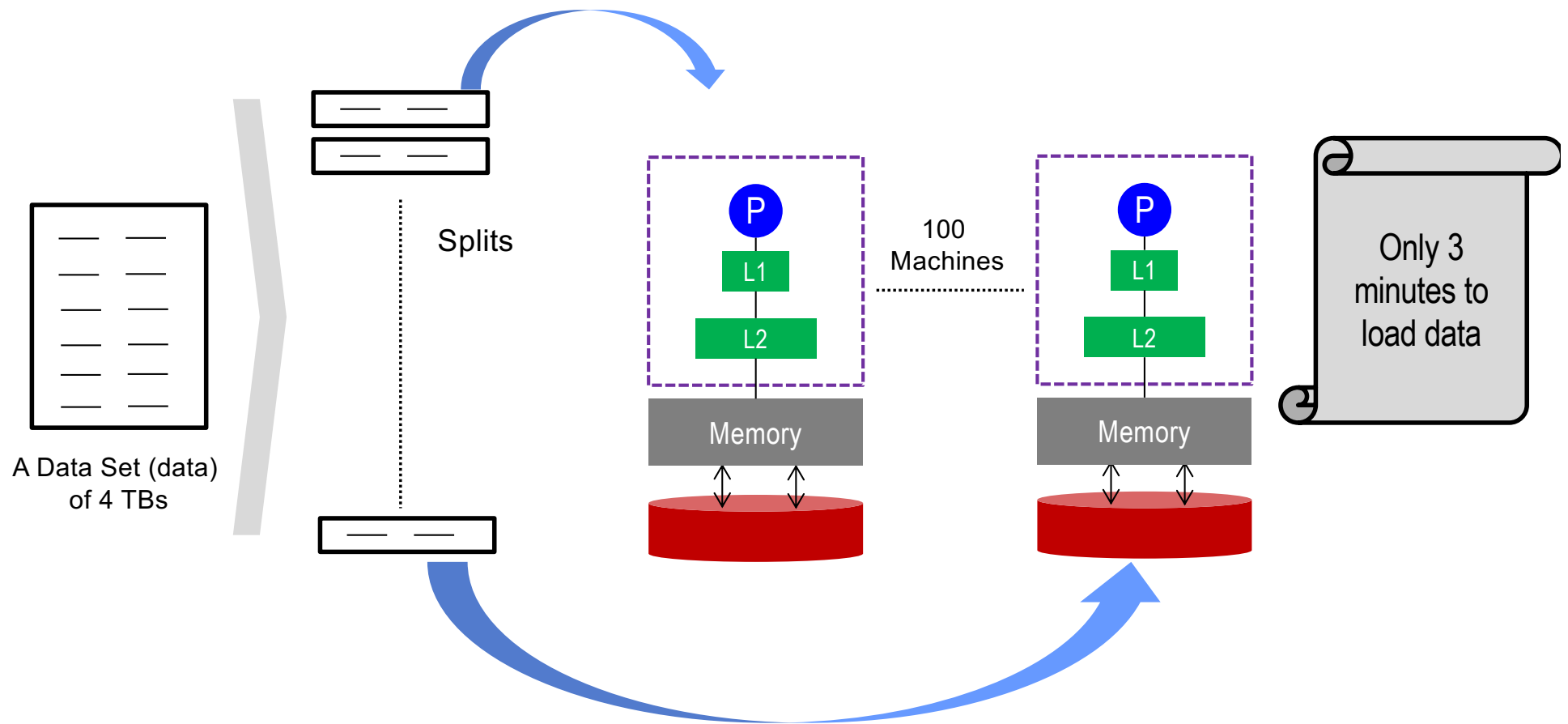


Why Distributed Systems?

- Even if 100s or 1000s of cores are placed on a CMP, it is a challenge to deliver input data to these cores fast enough for processing.



Why Distributed Systems?



Requirements

- But this requires:
 - A way to express the problem as parallel processes and execute them on different machines ([Programming Models and Concurrency](#)).
 - A way for processes on different machines to exchange information ([Communication](#)).
 - A way for processes to cooperate, synchronize with one another and agree on shared values ([Synchronization](#)).
 - A way to enhance reliability and improve performance ([Consistency and Replication](#)).

Requirements

- But this requires (*Cont.*):
 - A way to recover from partial failures ([Fault Tolerance](#)).
 - A way to secure communication and ensure that a process gets only those access rights it is entitled to ([Security](#)).

Text Books

- The primary textbooks for this course are:

1. Andrew S. Tannenbaum and Maarten Van Steen, *Distributed Systems: Principles and Paradigms*, 2nd E, Pearson, 2007.
2. George Coulouris, Jean Dollimore, Tim Kindberg, and Gordon Blair, *Distributed Systems: Concepts and Design*, 5th E, Addison Wesley, 2011
3. James E. Smith, and Ravi Nair, *Virtual Machines: Versatile Platforms for Systems and Processes*, 1st E, Morgan Kaufmann, 2005.

- Reference Book:

4. Tom white, *Hadoop: The Definitive Guide*, 2nd E, O'Reilly Media, 2011

email: chungdt@soict.hust.edu.vn

Q&A