CS 764: Topics in Database Management Systems

Lecture 1: Introduction

Xiangyao Yu

9/2/2020

Who am I?

Name: Xiangyao Yu

Assistant professor in computer sciences, database group

Research interests:

- Transaction processing
- New hardware for databases
- Cloud databases

Today’s Agenda

What is this course about?

Course logistics

Topics Covered in CS 764

- Query processing and buffer management
  - Join
  - Buffer management
  - Query optimization
- Advanced transaction processing
  - Granularity of locking
  - Optimistic concurrency control
  - B-tree
  - Aries recovery
  - Two-phase commit (2PC)
- Parallel and distributed DB
  - Parallel database
  - Distributed database
  - MapReduce
- **Guest lectures**

OLTP vs. OLAP

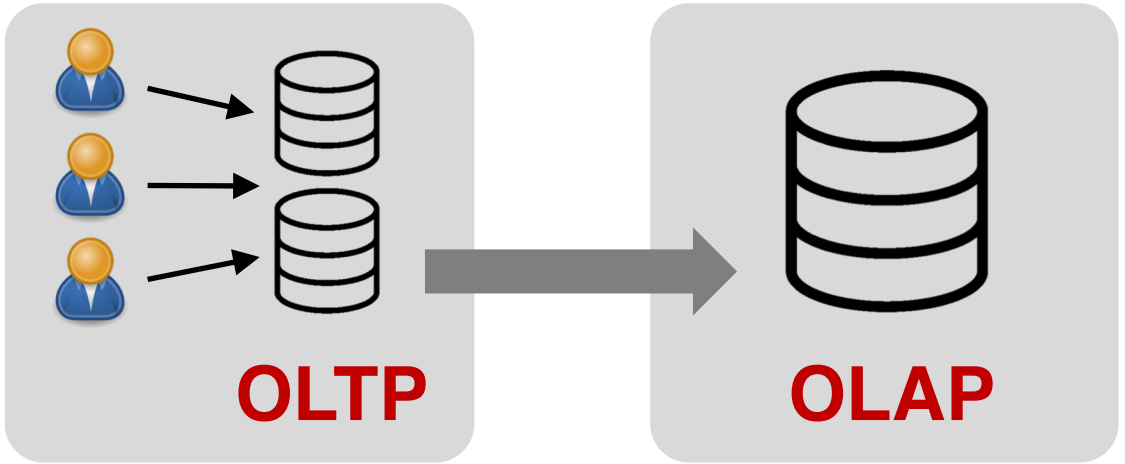
OLTP: On-Line Transaction Processing

- Users submit transactions that contain simple read/write operations
- Example: banking, online shopping, etc.

OLAP: On-Line Analytical Processing

- Complex analytics queries that reveal insights behind data
- Example: business report, marketing, forecasting, etc.

In modern databases, OLTP and OLAP are typically managed by two systems where the OLTP engine sends data to the OLAP engine periodically



Part I: Query processing and buffer management

Join (Lecture 1)

Table: **Orders**

OrderID	CID	OrderDate
10308	2	1996-09-18
10309	1	1996-09-19
10310	3	1996-09-20

Table: **Customers**

CID	CustomerName	ContactName	Country
1	Alfreds Futterkiste	Maria Anders	Germany
2	Ana Trujillo helados	Ana Trujillo	Mexico
3	Antonio Moreno Taquería	Antonio Moreno	Mexico

```
SELECT CustomerName, OrderDate
FROM Orders, Customers
WHERE Orders.CID = Customers.CID
```

Buffer Management (Lecture 2)

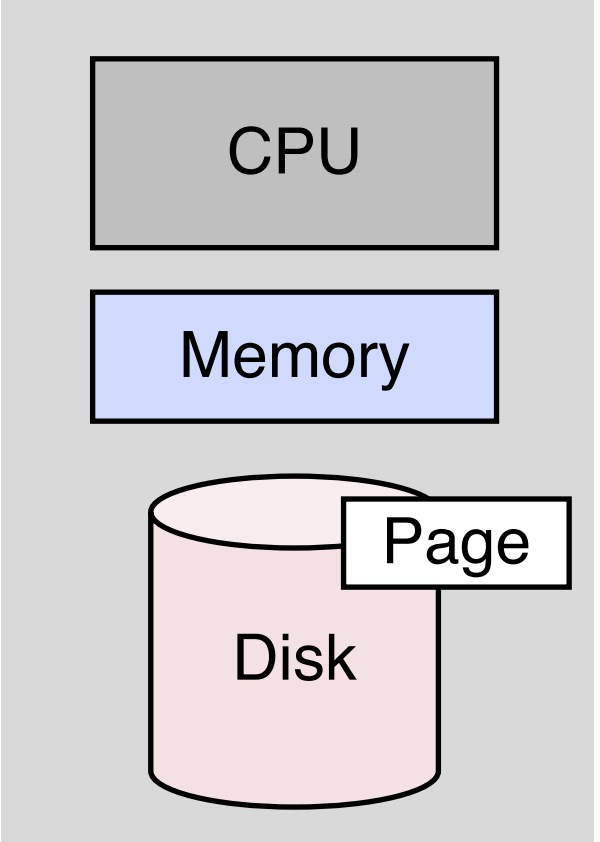
By default, data is stored on disk

Memory is orders of magnitude faster than disk

What pages to keep in memory is critical to performance

(Classic caching problem with its unique properties)

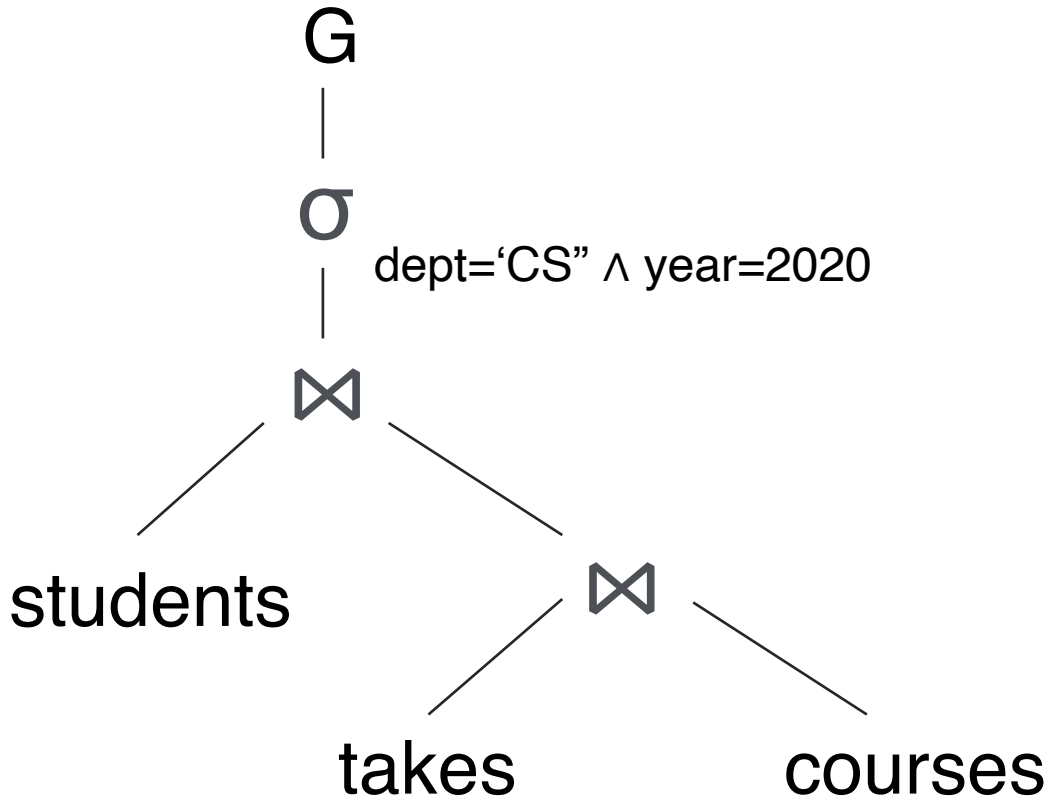
Transparent to higher level DB operations



Query Optimization (Lecture 3)

Tables: Course, Student, Takes

```
SELECT Course.name, count(*)
FROM Student, Course, Takes
WHERE Student.sid = Takes.sid
        Course.cid = Takes.cid
        Student.dept = 'CS'
        Student.year = 2020
GROUP BY Course.name
```



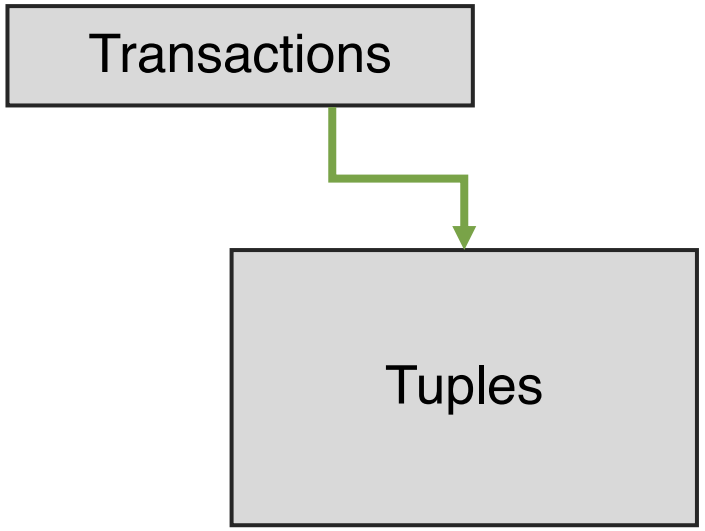
Is this plan optimal?

Part II: Advanced transaction processing

Transaction

The basic unit of work in a database

ACID: **A**tomicity, **C**onsistency, **I**solation, **D**urability

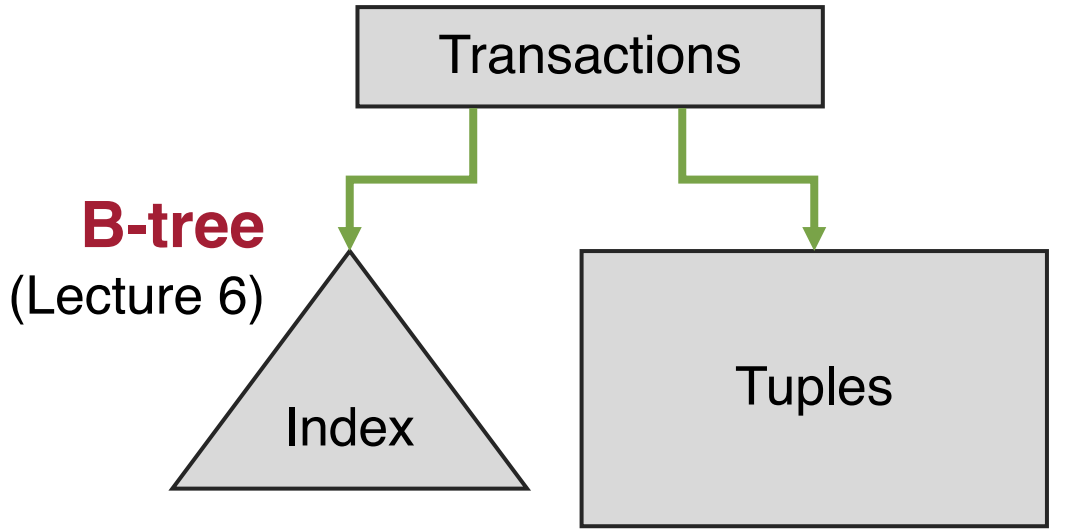


Concurrency control
(Lecture 4 & 5)

Transaction

The basic unit of work in a database

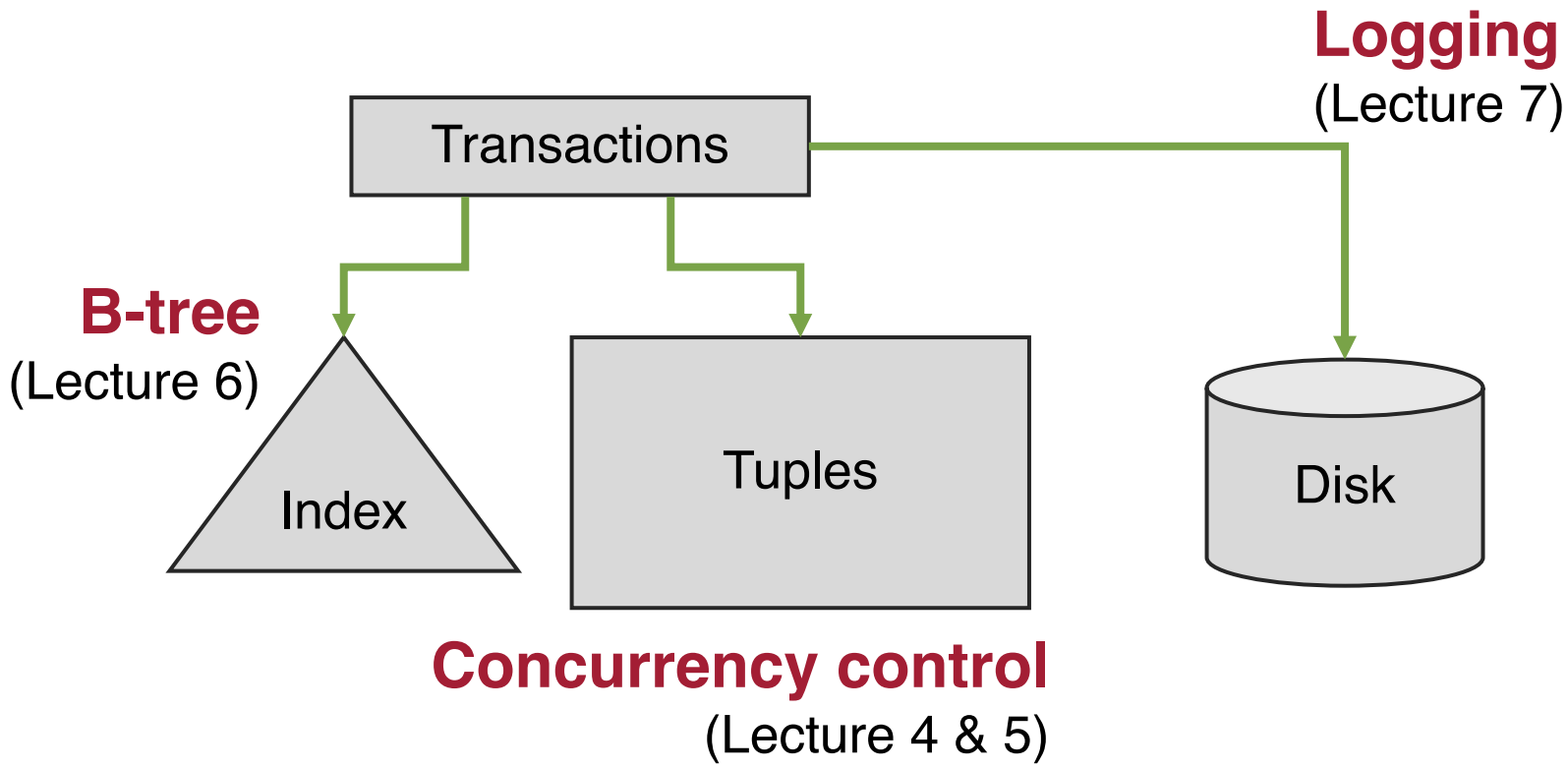
ACID: **A**tomicity, **C**onsistency, **I**solation, **D**urability



Transaction

The basic unit of work in a database

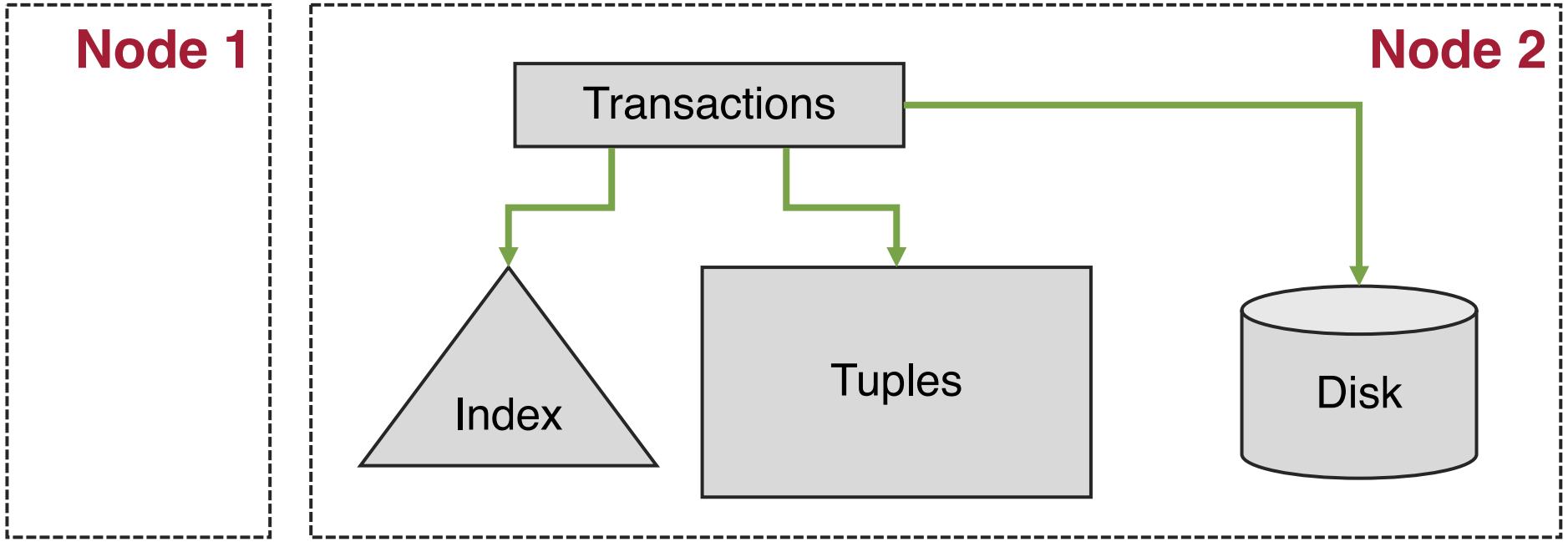
ACID: **A**tomicity, **C**onsistency, **I**solation, **D**urability



Transaction

The basic unit of work in a database

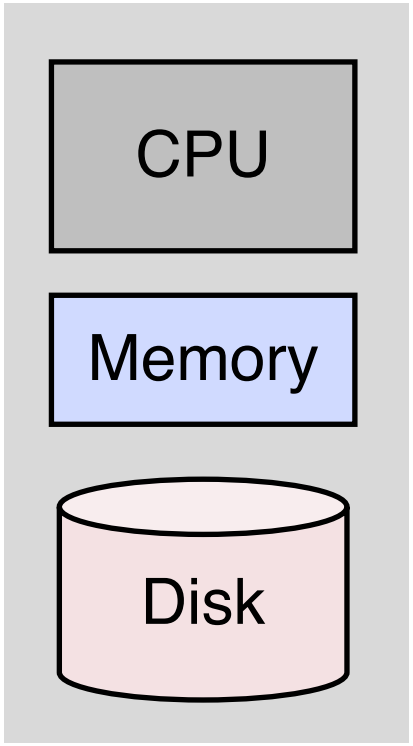
ACID: **A**tomicity, **C**onsistency, **I**solation, **D**urability



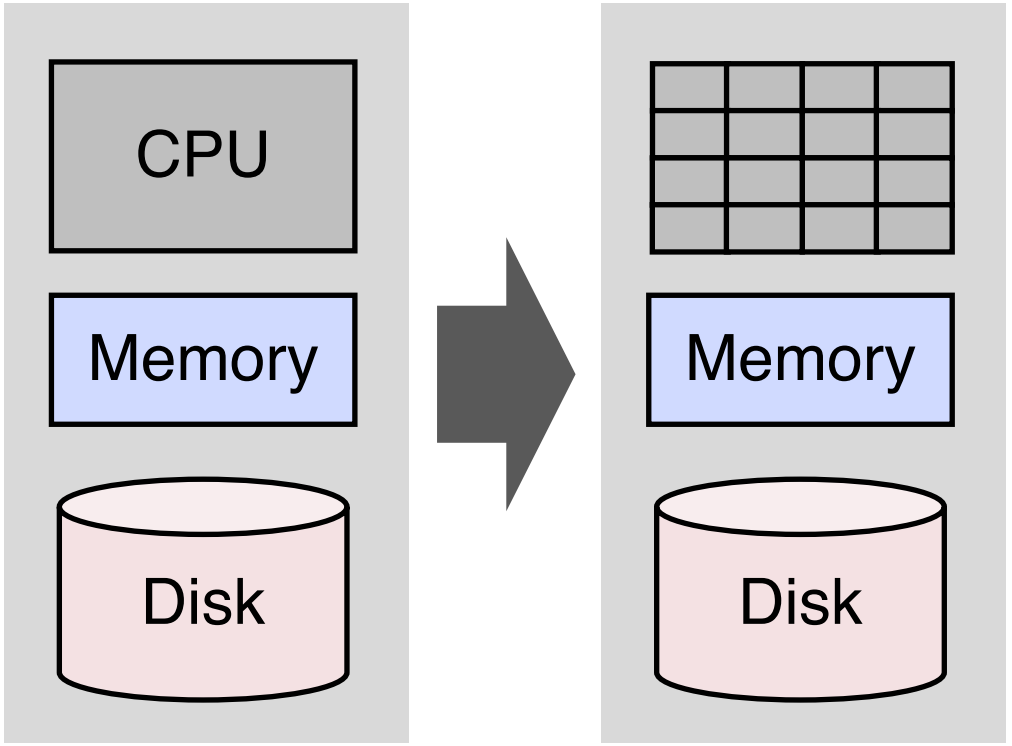
Two-Phase Commit (2PC) (Lecture 8)

Part III: Parallel and distributed DB

Parallelism

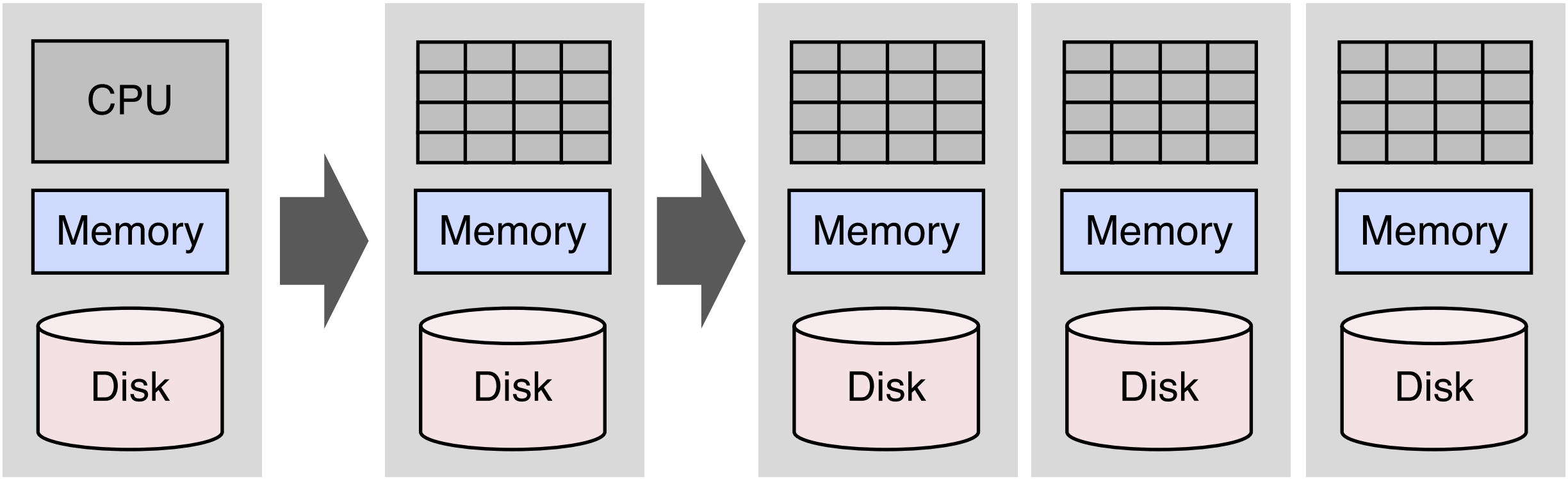


Parallelism



Multicore CPU
or GPU
(Lecture 9)

Parallelism



Multicore CPU
or GPU
(Lecture 9)

Distributed databases
(Lecture 10)

Course Logistics

Course Information

Website: pages.cs.wisc.edu/~yxy/cs764-f20

Prerequisite: CS 564

Office Hour: Monday 2:30—3:30pm on BBCollaborate Ultra

Reference textbooks:

- Red book
- Cow book

Lecture Format

Paper reading: one classic paper per lecture

Upload review: https://wisc-cs764-f20.hotcrp.com (please try to submit at least 60 min before the lecture starts)

- Paper summary
- Strength
- Weakness
- Comments and questions

Lecture

Group discussion (groups of 4—6 students)

Submit discussion summary: https://wisc-cs764-f20.hotcrp.com

Grading

Paper review: 15%

Exam: 35%

Project proposal: 5%

Project final report: 35%

Project presentation: 10%

Course Project

In **groups of 2—4** students

A list of example project ideas will be provided but you are encouraged to propose your own ideas

See previous projects here:

http://pages.cs.wisc.edu/~yxy/cs764-f20/dawn19.pdf

Computation Resources

CloudLab

https://www.cloudlab.us/signup.php?pid=NextGenDB

Chameleon

https://www.chameleoncloud.org

AWS: Apply for free credits at

https://aws.amazon.com/education/awseducate/

Oracle Cloud: Please contact the instructor is you are interested in this option.

Important Dates

Proposal due: Oct. 21

Exam: Nov. 4

Presentation: Dec. 7 & 9

Paper submission: Dec. 18

Before next lecture

Next lecture is Sep. 9, Wednesday

Read the following paper and submit review

- Leonard D. Shapiro, Join Processing in Database Systems with Large Main Memories. ACM Trans. Database Syst. 1986.

Register for https://wisc-cs764-f20.hotcrp.com