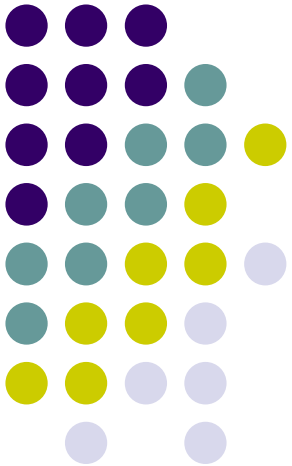
File Structures

Ch00. Introduction to Course



2020. Spring

Instructor: Joonho Kwon

jhkwon@pusan.ac.kr

Data Science Lab @ PNU

datalab

data science laboratory

Lecturer Information

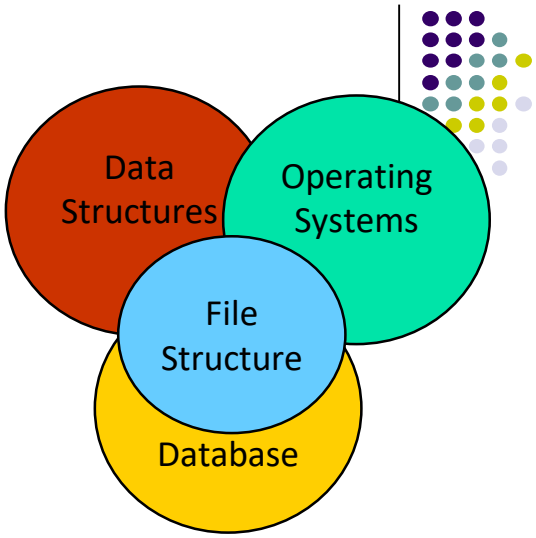
- **Name**:
  - 권준호 (Joonho Kwon)
- **Email:**
  - jhkwon@pusan.ac.kr
- **Website:**
  - https://datalab.pusan.ac.kr
- Research Topics
  - Big Data, IoT data management and analytics, Data Mining, Machine Learning, Data Science, Serious Games

Lecturer Information

- Office
  - 자연대연구실험동 416호 (Bldg. 313, Room 416)
  - Phone 051-510-3149

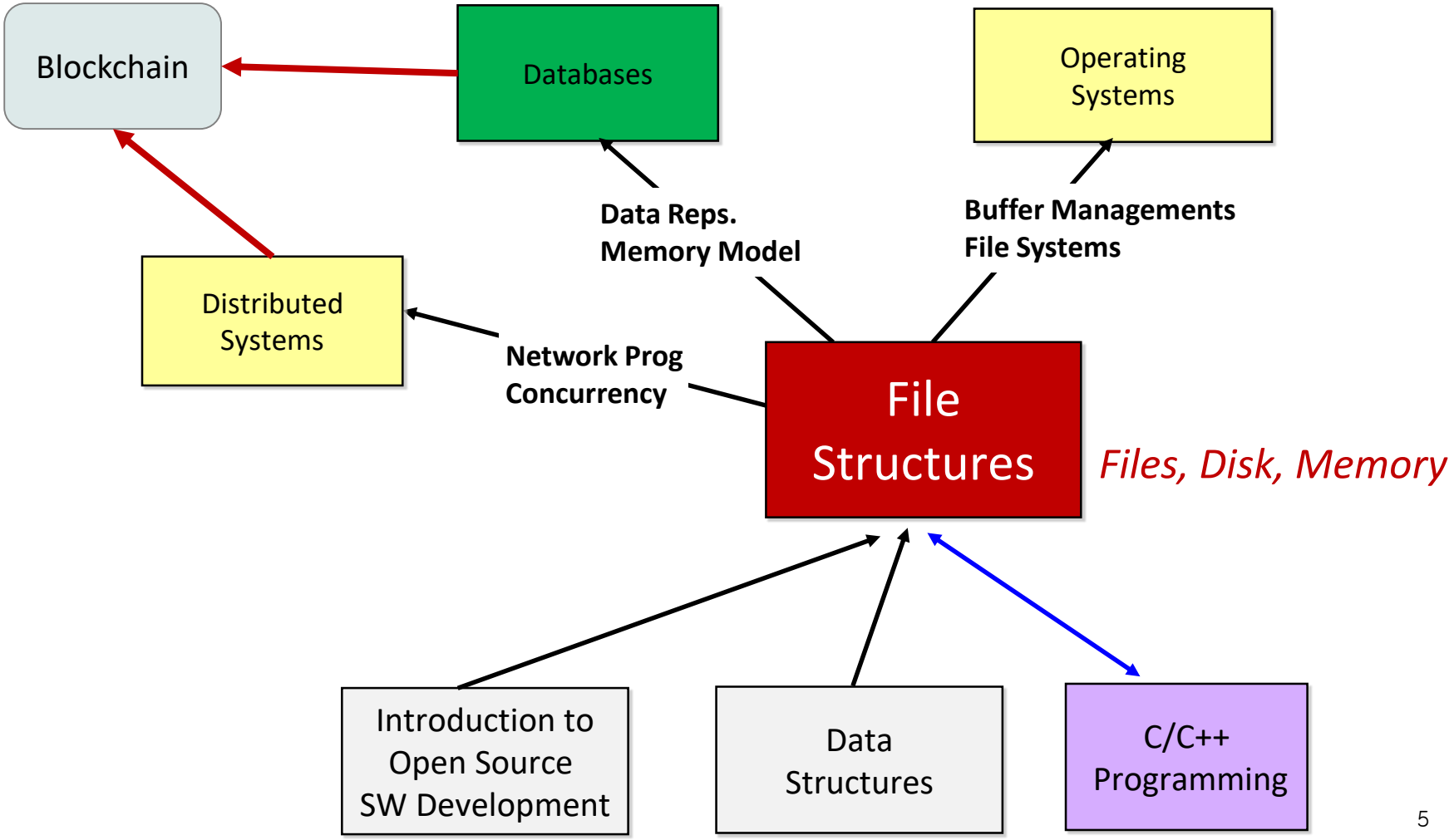
- Office Hour
  - Friday 14:00 -16:00

Course topic views



- **File structure overlaps with:**
  - **Data Structures**
    - In memory structure
      - List, Stack, Queue
      - Sorting: Heap, Merge sort
      - Tree: AVL tree, B-Tree
      - Hashing:
  - **Databases**
    - Large-scale data, File, Buffer, Query processing
    - Disk-based data structure
  - **Operating System**
    - Hard Disk Drives, RAID, Flash-based SSDs
    - Files and Directories, File System Implementation

Role within CSE Curriculum



Topics and course Schedule (1/2)

- Tentative Schedule

Week No.	Topics and Activities
Week 1	Introduction to the Design and Specification of File Structures
Week 2	Fundamental File Processing Operations
Week 3	Database Architecture
Week 4	Organization of Disks
Week 5	Fundamental File Structure Concepts
Week 6	Managing Files of Records
Week 7	Organizing Files for Performance
Week 8	**Midterm**

Topics and course Schedule (2/2)

Week No.	Topics and Activities
Week 9	Indexing
Week 10	Multi-level indexing
Week 11	B-tree
Week 12	Indexed Sequential File Access and Prefix B Trees
Week 13	Hashing
Week 14	Extendible Hashing
Week 15	**Final Exam**
Week 16	

Required Skills

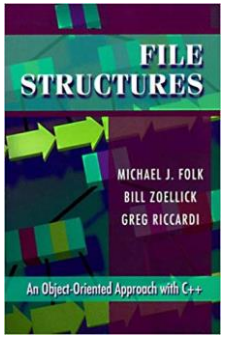
- C++ Programming
  - Inheritance, template
- How to compile c++ program in Linux or Windows
  - g++, make, gnu debugger (used in the class)
- Source code management
  - git, github

Course Website

- Course Website
  - https://plato.pusan.ac.kr/course/view.php?id=87239

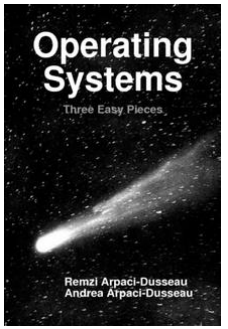
- Students should check the website regularly
  - All course materials will be provided
  - Notices
  - Assignments
  - Projects

Textbooks (1/3)



- Required
  - **File Structures, An Object-Oriented Approach with C++**

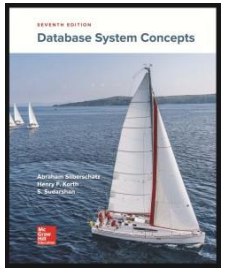
    Third Edition, Addison Wesley, 1997
  - This book really matters for the course!
    - How to solve labs
    - Practice problems typical of exam problems



  - Operating Systems: Three Easy Pieces
    - http://pages.cs.wisc.edu/~remzi/OSTEP/
      - Free download
    - Some chapters
      - Hard Disk Drives, RAID, Flash-based SSDs
      - Files and Directories, File System Implementation

Textbooks (2/3)

- Optional
  - Database books (any book is okay)
    - Database System Implementation
    - Database System concepts

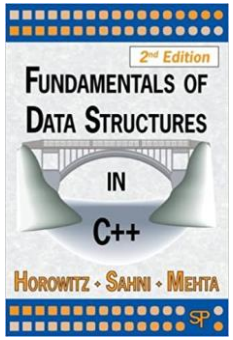


  - **화일구조 (C++을 사용한 객체 지향접근 방식) (Korean translation)**

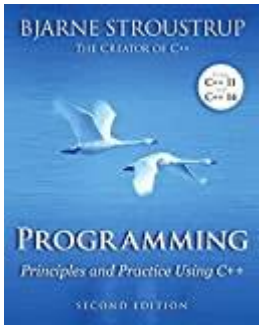
  - (Optional) 화일구조 (another Korean book)
    - 이석호 저, 정익사

Textbooks (3/3)

- Optional
  - Fundamentals of Data Structures in C++



  - The C++ programming language
    - http://www.stroustrup.com/
      - You can watch several videos



Bjarne Stroustrup pronounce **c++ as c plus plus**

Language for the class

- **Only c++ can be used**

- If you do not have any knowledge for c++ languages, please drop this course

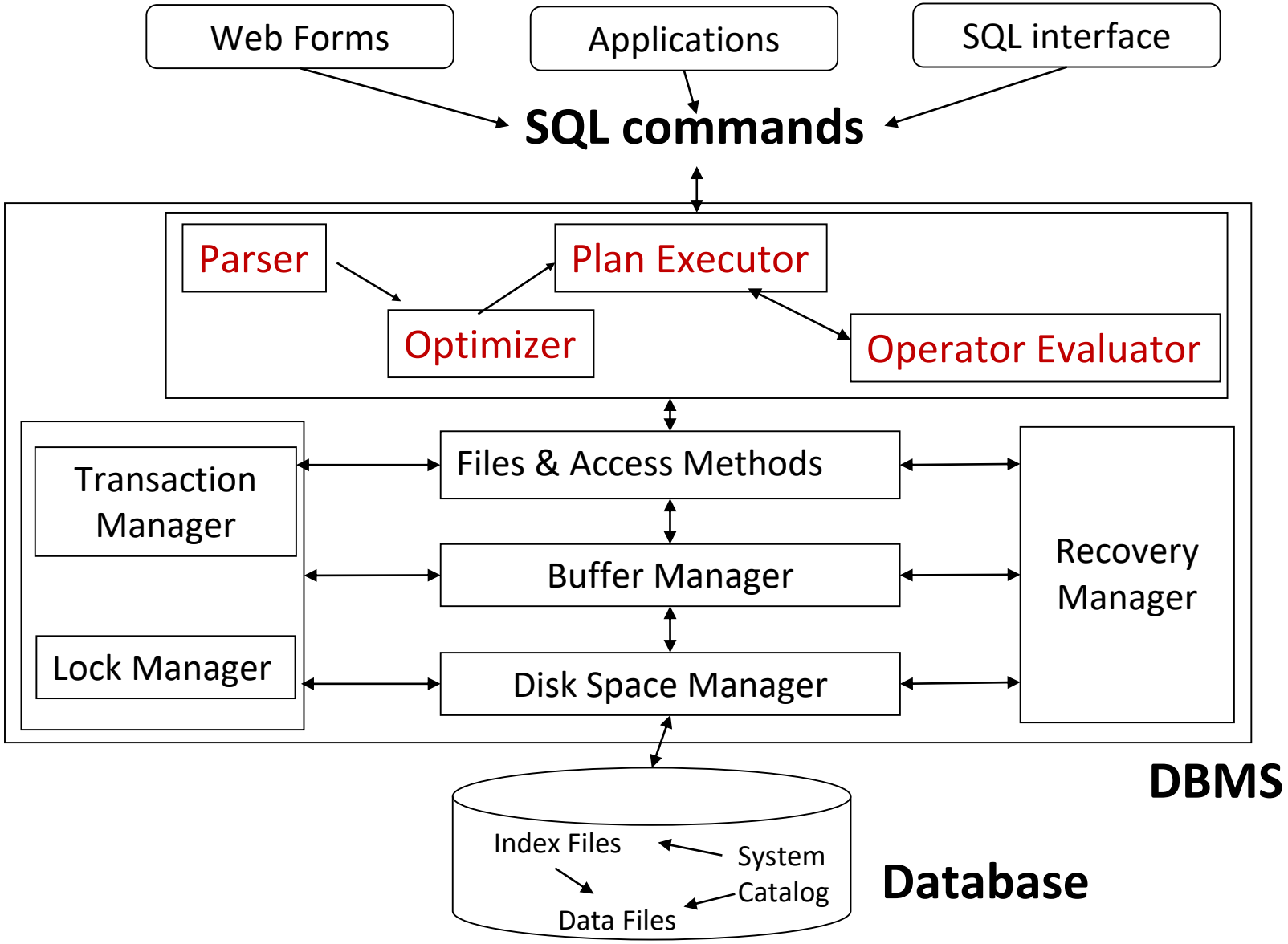
About class (1/2)

- Grading policy
  - Midterm: 30%
  - Final: 30%
  - Homework: 37%
  - Etc: 3%

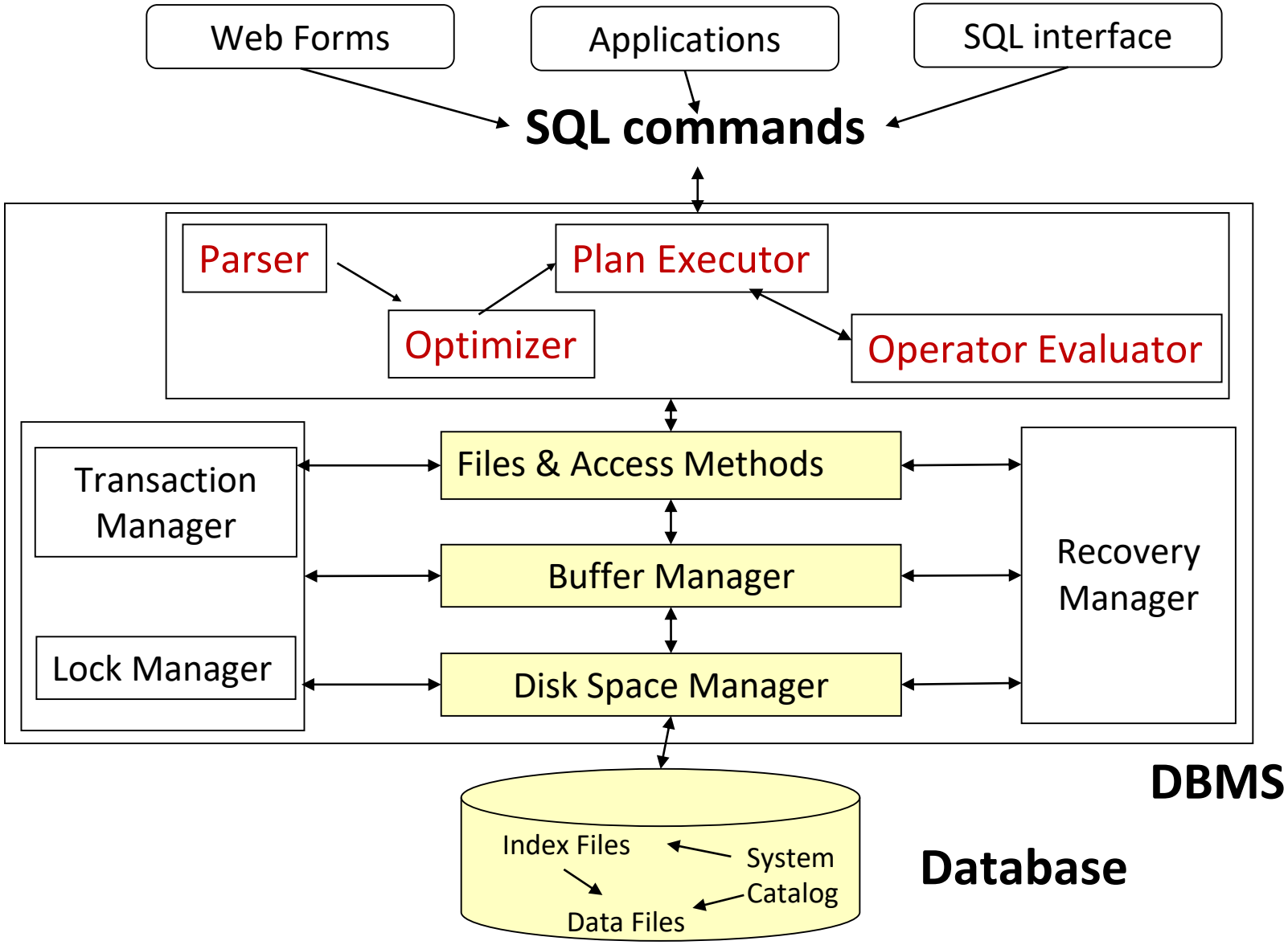
About class (2/2)

- Note!!!
  - If a student does not take either a midterm or a final exam without any notices before the exams, he/she will get a F grade (no exceptions !!)
  - If a student submit homework less than half, he/she will get a F grade (no exceptions !!)

Big Picture: Architecture of a DBMS



The focus of file this course course



Q&A