

2/4 B.Tech. SECOND SEMESTER

File Structures

CS4T4

Required

Credits: 4

Lecture: 4 periods/week

Internal assessment: 30 marks

Tutorial: 1 period /week

Semester end examination: 70 marks

Course context and Overview: This subject provides basics underlying architecture of databases. It explains field and record formation along with effective searching process by using B Trees and B+ Trees.

Prerequisites: Data Structures

Objectives:

1. Provide an introduction to the fundamental file operations and storage systems.
2. Introducing fundamental concepts of file structure.
3. Introducing the most important high-level file structures tools which include indexing, co sequential processing, B trees, Hashing.
4. Applying the techniques in the design of C++ programs for solving various file management problems.

Learning Outcomes:

Ability to:

1. Understand the fundamental concepts of file processing operations and storage structures.
2. Apply the object orientation concepts to manipulate records.
3. Apply the concepts of sorting and merging on multiple files.
4. Analyze the sequential and indexing file accessing techniques with appropriate data structures.
5. Illustrate the usage of hashing techniques to organize file structures.

Unit I

Fundamental File Processing Operations:

Physical Files and Logical Files, Opening Files, Closing Files, Reading and Writing, Seeking, Special Characters, The Unix Directory Structure, Physical devices and Logical Files. Primary file organization: Sequential file organization, Direct file organization.

Unit II

Secondary Storage and System Software:

Disks, Magnetic Tape, Disk versus Tape; CD-ROM: Introduction, Physical Organization, Strengths and Weaknesses; Storage as Hierarchy, A journey of a Byte, Buffer Management, Input /Output in UNIX.

Unit III

Fundamental File structure Concepts:Field and Record Organization, Using Classes to Manipulate Buffers, Using Inheritance for Record Buffer Classes, Managing Fixed Length, Fixed Field Buffers

Managing files of records: Record Access, More about Record Structures, Encapsulating Record Operations in a Single Class, File Access and File Organization.

Unit IV

Consequential Processing A Model for Implementing Consequential Processes, Application of the Model to a General Ledger Program, Extension of the Model to include Multiway Merging.

Sorting of large Files: A Second Look at Sorting in Memory, Merging as a Way of Sorting Large Files on Disk, sorting files on Tape, sorting and merging in unix

Unit V

Indexing: what is an Index?, A simple Index for entry -sequenced files, Indexes that are too large to hold in memory, Indexing to provide access by multiple keys, Retrieval using combinations of secondary key, Improving the secondary index structure, binding.

Multilevel Indexing and B-Trees: The invention of B-Tree, Statement of the problem, Indexing with Binary Search Trees; Multi-Level Indexing, B-Trees, Example of Creating a B-Tree, B-Tree Methods; Nomenclature, Formal Definition of B-Tree Properties, Deletion, Merging and redistribution.

Unit VI

Indexed Sequential File Access And Prefix B + Trees:

Indexed Sequential Access, Maintaining a Sequence Set, Adding a Simple Index to the Sequence Set, The Content of the Index: Separators Instead of Keys, The Simple Prefix B+ Tree and its maintenance, Index Set Block Size, Internal Structure of Index Set Blocks: A Variable-order B- Tree, Loading a Simple Prefix B+ Trees, B-Trees, B+ Trees and Simple Prefix B+ Trees in Perspective.

Unit VII

Hashing: Introduction, A Simple Hashing Algorithm, Hashing Functions and Record Distribution, How much Extra Memory should be used?, Collision resolution by progressive overflow, Buckets, Making deletions, Other collision resolution techniques, Patterns of record access.

Unit VIII

Extendible Hashing: How Extendible Hashing Works, Implementation, Deletion, Extendible Hashing Performance, Alternative Approaches.

Learning Resources

Text Books:

1. File Structures :An object-oriented approach with c++, Michael Folk, Bill Zoellick, Grg Riccardi, Pearson Education.

Reference:

1. Data Management and File structures, Mary E.S. Loomis ,2nd Edition, PHI
2. File Organization and Processing , Alan I. Tharp ,Wiley indian edition.