B-trees

Instructor: Thanh-Chung Dao
(chungdt@soict.hust.edu.vn)
Slides by Dr. Ta Tuan Anh

1

Vấn đề

- Cần lưu trữ số phần tử dữ liệu rất lớn
- Lưu trữ trên bộ nhớ ngoài
- Tìm kiếm nhanh
- Phân trang dữ liệu
  - Tăng số nhánh cây => giảm chiều cao cây
  - Gom nhóm dữ liệu thành những block, giảm số lần truy xuất trên đĩa
- Dùng B-Tree

2

Định nghĩa

Một B-Tree bậc n có các tính chất (1972)
- Mỗi nút có tối đa 2n khóa
- Mỗi nút (không là nút gốc) có ít nhất là n khóa
- Mỗi nút hoặc là nút lá hoặc có tối thiểu n+1 node con
- Cây Bậc 2
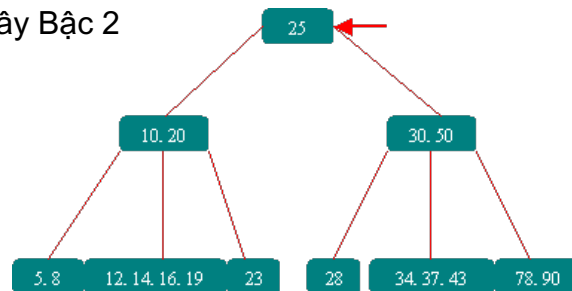
3

B-Tree in the wild

- Red-black trees: widely used as system symbol tables
  - Java: java.util.TreeMap, java.util.TreeSet.
  - C++ STL: map, multimap, multiset.
  - Linux kernel: linux/rbtree.h.
- B-Trees: widely used for file systems and databases
  - Windows: HPFS.
  - Mac: HFS, HFS+.
  - Linux: ReiserFS, XFS, Ext3FS, JFS.
  - Databases: ORACLE, DB2, INGRES, SQL, PostgreSQL

- All nodes in B-Tree are assumed to be stored in secondary storage (disk) rather than primary storage (memory),
- There basic operations for accessing a page: *Disk-Read(), Disk-Write(), Allocate-Node()*

4

B-Tree Library

- Software and documentation is accessed at http://www.hydrus.org.uk/doc/bt/html/index.html
- https://hydrus.org.uk/downloads/bt-5.0.0.tar.gz

5

API

- Creating a B Tree File

BTA* btcrt(char* *fid*, int *nkeys*, int *shared*);

- Opening a B Tree File

BTA* btopn(char* *fid*, int *mode*, int *shared*);

- Closing a B Tree File

int btcls(BTA* *btact*);

6

API (cont.)

- Inserting a key and data

int btins(BTA* *btact*, char* *key*, char* *data*, int *dsize*);

- Updating data for an existing key

int btupd(BTA* *btact*, char* *key*, char* *data*, int *dsize*);

- Locating data for an existing key

int btsel(BTA* *btact*, char* *key*, char* *data*, int *dsize*, int* *rsize*);

- Deleting a key and associated data

int btdel(BTA* *btact*, char* *key*);

- Locating data for the next key in sequence

int btseln(BTA* *btact*, char* *key*, char* *data*, int *dsize*, int* *rsize*);

7

Building and installing the BT Library

- Unpack the tar file into a convenient directory.

$cd <bt library>

$make clean

$make

- Make built an UNIX static library **libbt.a**, a BT test harness **bt**, and a utility, **kcp**, which performs intelligent copies of BT index files.

8

Quiz 1

- Install and compile BT Library in your machine
- Run BT test harness to verify if successful installed
- See documentation at
https://hydrus.org.uk/doc/bt/html/ch05.html

9

Quiz 2

- Use the BT library to write a phone book program that manipulates data on the secondary disk.

10

Another library for B-Tree

- Download at
  http://www.mycplus.com/utilitiesdetail.asp?iPro=10

- This library allows specifying different comparison functions for keys.

11

Mini project 1

- Make a program to manage a computer dictionary
  - Add/Search/Delete a word (using B-Tree)
  - Auto complete search. Ex. When we enter "comput" and <tab>, the word "computer" should be auto completed (like in Bash Shell)
  - Suggestion search => Use soundex library
  - GUI interface is optional, but will be a plus point
    - GTK
- Please test the performance of your program with a dictionary of millions words (the words can be randomly created)
  - Test for the two basic operations: search and insert
- Project in group of 3 students

12