CSE 373: Data Structures and Algorithms

Lecture 24: B-Trees
Disk Based Data Structures

• All data structures we have examined are limited to main memory
  – Have been assuming data fits into main memory

• Counter-example: transaction data of a bank > 1 GB per day
  – uses secondary storage (hard disks, tapes, etc)
  – operations: insert, delete, searches

• Idea: Make a search tree that is secondary storage enabled
Memory Hierarchy

• Big-Oh assumes all operations take the same amount of time
  – Is this really true?

• Cycle – time it takes to execute an instruction
Cycles to access:

- Registers: 1
- Cache: tens
- Main memory: hundreds
- Disk: millions
Hard Disks

• Large amount of storage but slow access
• Identifying a page takes a long time
  – Pays to read or write data in pages (i.e. blocks) of 0.5 – 8 KB in size