ICS 321 Fall 2009
A Toy Course Project

Asst. Prof. Lipyeow Lim
Information & Computer Science Department
University of Hawaii at Manoa
Book Catalog for Internet Store

• Requirements:
  – Read-intensive workload.
  – Many concurrent users.
  – Reasonably responsive browsing experience

• Query features:
  – Keyword search
  – Multi-faceted browsing
  – Hierarchical topic navigation

• Measures:
  – Latency of processing the various query features
  – Throughput: number of concurrent users

• Cost?
Possible Questions for Investigation

• DBMS vs file-system + text indexer?
• Using a DBMS what are the different ways of supporting
  – keyword search,
  – multi-faceted search,
  – hierarchical navigation?
• What table/schema design is needed for multi-faceted search?
• MySQL: which storage engine is suitable?

Too many questions?
Just pick a couple to tackle for the project!
DBMS vs FS+Indexer

• Pick a DBMS software. Say DB2.
• Pick an FS+indexer. Say Win FS + Apache Lucene
• Implement the required query features for these two setups.
• Setup a testbed
  -- Get some real data or generate synthetic and load the data into our setup
  -- Get a “driver” to simulate users browsing books, i.e. a multi-threaded program that fires queries against our system.
  -- Instrument a way to measure the latency and throughput.
• Run the driver and collect measurements over the two setups possibly over different parameters.
Analyze the results

• Analyze the results
• If something is not right,
  – investigate,
  – tweak your setup,
  – rerun,
  – re-analyze.
• Based on your results what are your conclusions and recommendations?

9/8/2009 Lipyeow Lim -- University of Hawaii at Manoa