CS425 Course Information

- Course webpage: www.cs.bilkent.edu.tr/~mustafa.ozdal/cs425
- Spare hour: Monday 8:40-9:30 (will be used starting from Oct 3 until further notice)

- Grading:
  - Midterm: 30%
  - Project: 30%
  - Attendance: 10%
  - Final: 30%

- FZ Policy: You should get 30 out of 70 (midterm + project + attendance)
Warm-up Puzzle

• Assume that you’re running a web server.
• You want to count the number of unique users that visit your home page.

• Details:
  • User IDs are 64-bit integers
  • There are at most U unique users in the world (U < 2^{64})
  • You have a continuous stream of users

• Which data structure/algorithm would you use?
• What is the space complexity?
Warm-up Puzzle (cont’d)

• Now assume that you’re running a web server at Amazon.
• You want to estimate the number of unique visitors for each page. *(There are millions of pages.)*

• What is the space complexity of the previous algorithm?

• Can you come up with an estimator that uses only a 64-bit integer as storage? i.e. The total memory used is only 64 bits per web page.
  • *Hint 1: Your estimator doesn’t have to be perfect, but should have reasonable accuracy when multiple of them are combined together.*
  • *Hint 2: Assume that you have a perfect hash function*