Today’s Agenda

0. Announcements
1. Calendar
2. Professor Ruigang Yang, UK Computer Science
3. Teem Geek Chapter 2
4. *The Magic of Computer Science*
   A. Algorithms and Structure
   B. Massive Parallelism (map reduce)
0. Announcements

• 13 October: Group Project Launch
  – Expect Group assignments and project details
• 20 October: Quiz 1
  – Begin to organize your notes for Quiz Study
• Office hours
  – Can check with Amy or Diane in CS department to see if a meeting and/or trip has curtailed office hours
1. Calendar

- HW3 ("numbers") due tonight by midnight tonight

```cpp
int main () {
    int SHIFT=10;  // number of shifts to print in table

    // ASCII codes to "decode" (this was given in the assigned problem)
    char A[55] = {80, 94, 25, 58, 107, 94, 25, 80, 98, 101, 93, 92, 90, 109, 108, 26};
    // Construct a string from the declared array of characters
    string message(A);

    // Print the table for each shift value, starting at zero
    for(int i =0; i < SHIFT; i++) {
        printShifted(message, -i);
        cout << endl;
    }
    return 0;
}
```
// Shifts each ASCII character in string by "shift"
void printShifted (string s, int shift) {
   // print shifted values as characters, fixed field width so that
   // it lines up
   for (int i=0; i < (s.length()); i++)
      cout << setw(3) << (char)(s[i]-shift);
   cout << endl;

   // print shifted values as ASCII equivalents
   for (int i=0; i < (s.length()); i++)
      cout << setw(3) << s[i]-shift;
   cout << endl;
}

2. Professor Yang

- PhD UNC Chapel Hill
- Full Professor at UK
Teem Geek

• Clickers
The Magic of Computer Science (Part 2)

Any sufficiently advanced technology is indistinguishable from magic

- Arthur C. Clarke
There is no magic.

- Ken Calvert
## Review: Powers of 2

<table>
<thead>
<tr>
<th>$2^0$</th>
<th>$2^1$</th>
<th>$2^2$</th>
<th>$2^3$</th>
<th>$2^4$</th>
<th>$2^5$</th>
<th>$2^6$</th>
<th>$2^7$</th>
<th>$2^8$</th>
<th>$2^9$</th>
<th>$2^{10}$</th>
<th>$2^{11}$</th>
<th>$2^{12}$</th>
<th>$2^{13}$</th>
<th>$2^{14}$</th>
<th>$2^{15}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>4</td>
<td>8</td>
<td>16</td>
<td>32</td>
<td>64</td>
<td>128</td>
<td>256</td>
<td>512</td>
<td>1024</td>
<td>2048</td>
<td>4096</td>
<td>8192</td>
<td>16384</td>
<td>32768</td>
</tr>
</tbody>
</table>
Recap

• The power of computing comes from the ability to manipulate digital information

• Analog-to-Digital conversion translates natural phenomena (e.g., waveforms) into digital representation
  – Music (MP3), voice telephony
  – Images (JPEG, HDTV)

• Once information is encoded digitally, it can be:
  – Stored on any digital medium
  – Manipulated by any computer
  – Duplicated perfectly at very low cost
Algorithms

Exploit structure of a problem / ordering

Very common problem:
    Order a set of information
Resumes

• Last Name, First Name
• Lots of other info...

• Problem – return resumes to owners (after ingesting information, of course)
Ordering is Crucial for Queries

• Queries (or “data mining”)
  – Find information in the set of data to answer questions
  – Information can be composite:
    • How many people are from Jessamine county?
  – Or individual
    • What is Jamie Lewis’ age?

• Queries are *hard* without structure / ordering
Algorithm for Ordering

• Multiple stages
  – “sort” smallest stack
  – Reduce number of stacks via merge
Ordering a.k.a. Sorting

• Sort
  – Sort a single group of papers any way you want. Make sure at the end your group of papers is in alphabetical order by last name.
  – At the end of this phase we should have a bunch of stacks of papers, each of which is alphabetized.
Total Ordering: Can Merge to Reduce Number of Stacks

• Merger task: merge two stacks together, keeping them in alphabetical order
• Now find another merger and negotiate again (one of you drops out, the other continues to the next round as a merger)
• Double down!
Alternate Plan: Shards

• How about creating shards based on first letter of last name?
• 26 shards
• Now order each shard

• Any relationship to binary numbers?

• There is no magic.
Ordered Data: Can Answer Queries

- Ordered / structured data very powerful
- But there is just one of me to rifle through the data source to answer questions....

- What about information like this?
  - What are the top Kentucky counties (by count) represented in this class?
Parallelism

• Structure of many problems allows the possibility to solve them in pieces, where the pieces happen simultaneously

(There is no magic)
MapReduce

• Programming model for solving problems
  – In parallel
  – On a cluster
  – Low level abstractions (binary numbers)
  – Structure of information (algorithms)
  – Massively parallel / distributed solutions driven by data centers available on a global scale
TakeAways

• Numbers are an abstraction
  – At the lowest level, they are binary sequences

• Media is an abstraction
  – At the lowest level, digital media is a binary sequence

• Algorithms are abstractions
  – At the lowest level, they are sequences of operations on binary sequences

• Parallelism is an abstraction
  – At the lowest level, it is simultaneous, structured activity on binary sequences
Distributing Results

• Collect and show results