Let's Talk Storage

<table>
<thead>
<tr>
<th></th>
<th>#Articles</th>
<th>Storage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Encyclopedia Britannica</td>
<td>40,000+</td>
<td>32 Heavy Books</td>
</tr>
<tr>
<td>Microsoft Encarta</td>
<td>60,000+</td>
<td>1.5 Petabytes</td>
</tr>
<tr>
<td>Wikipedia</td>
<td>4,723,819 (in English)</td>
<td>5 TB</td>
</tr>
</tbody>
</table>

How can Google process so much information so fast?

Now, we have some complex calculations to do.

What problems did we hit?
- Communication
- Time?
- Calculations, load, and correctness

How could we optimize the process?

What things can't we prevent?

average 1,000,000 numbers
1 CPU
250,000 to each
average the averages
Token Ring
Amdahl's Law

\[ \text{Speedup} = \frac{1}{(1-P) + \frac{P}{N}} \]

- \( P \) = fraction of program that is parallel
- \( N \) = number of processing entities

Problems that distributed systems face:

1. **Election (Ring Alg)**
2. **Bully Alg**
3. Send node to everyone w/ highest ID
4. If you get in charge from machine and you \( i \) > \( n \)
5. Last one standing gets token