

Hardware and Data

LBSC 690: Jordan Boyd-Graber

September 10, 2012



COLLEGE OF
INFORMATION
STUDIES

Adapted from Jimmy Lin's Slides

Outline

- 1 History
- 2 Introductions
- 3 Computing
- 4 Networking
- 5 How Big and How Fast?
- 6 How Fast?

A COMPUTER WANTED.

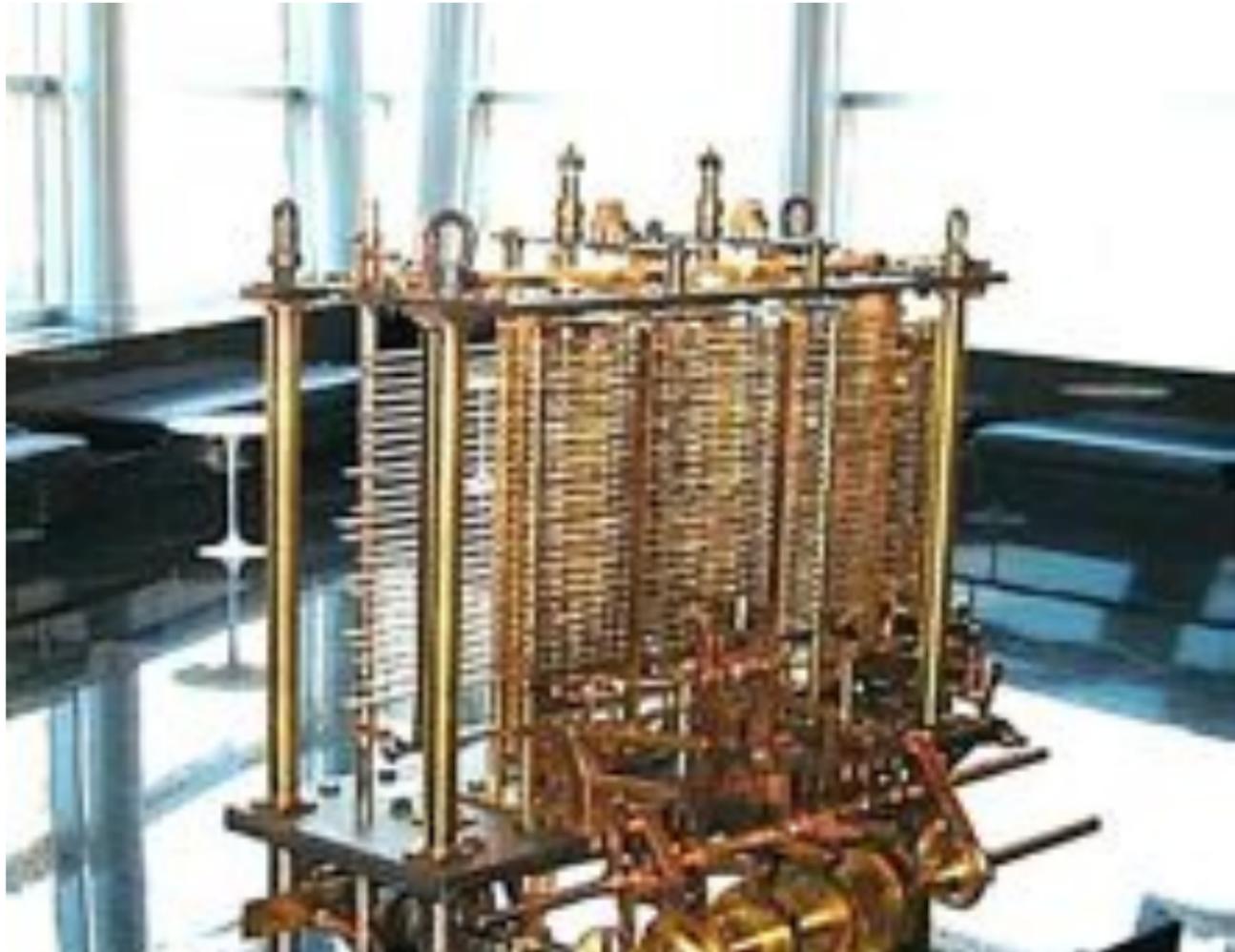
WASHINGTON, May 1.—A civil service examination will be held May 18 in Washington, and, if necessary, in other cities, to secure eligibles for the position of computer in the Nautical Almanac Office, where two vacancies exist—one at \$1,000, the other at \$1,400..

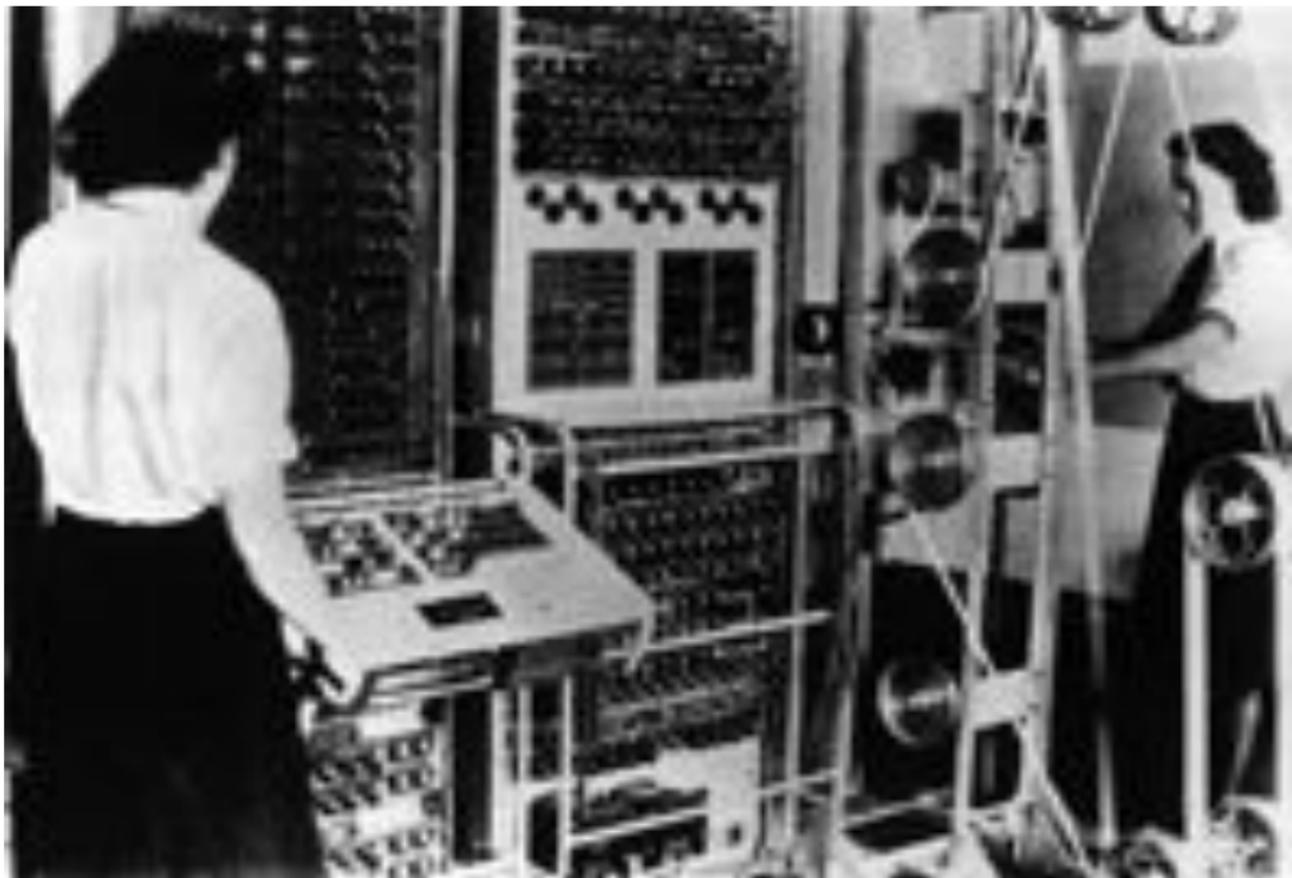
The examination will include the subjects of algebra, geometry, trigonometry, and astronomy. Application blanks may be obtained of the United States Civil Service Commission.

The New York Times

Published: May 2, 1892

Copyright © The New York Times















Looking at a Webpage

- How is it stored on the server
- How does the information reach your computer
- How do people interact with the webpage
- How do you create a webpage
- How can the webpage interact with large amounts of data
- How can you find the webpage in the Internet
- How can you put multimedia on the webpage
- How can you get computers in front of people

Looking at a Webpage

- How is it stored on the server (Week 1)
- How does the information reach your computer (Week 2)
- How do people interact with the webpage (Week 3)
- How do you create a webpage (Week 4-5)
- How can the webpage interact with large amounts of data (Week 7)
- How can you find the webpage in the Internet (Week 8)
- How can you put multimedia on the webpage (Week 9)
- How can you get computers in front of people (Week 10)

Looking at a Webpage

- How is it stored on the server
Hardware - Software
- How does the information reach your computer
Computer - Computer
- How do people interact with the webpage
Computer - Human
- How do you create a webpage
Human - Computer
- How can the webpage interact with large amounts of data
Data - Computer
- How can you find the webpage in the Internet
Computer - Computer - Human
- How can you put multimedia on the webpage
Artist - Viewer
- How can you get computers in front of people
Money - Consumer

Outline

- 1 History
- 2 Introductions**
- 3 Computing
- 4 Networking
- 5 How Big and How Fast?
- 6 How Fast?

- Second year assistant professor
 - ▶ iSchool and UMIACS
 - ▶ Offices: 2118C Hornbake / 3126 AV Williams
- Second time teaching the class
- Born in Colorado
- Grew up in Iowa (home)
- Went to high school in Arkansas
- Undergrad in California
- Grad school in New Jersey
- Brief jobs in between:
 - ▶ Working on electronic dictionary in Berlin
 - ▶ Worked on Google Books in New York
- ying / jbg / jordan / boyd-graber

This Course

- Mailing list
- Office hours
- Discussion Points
- Note cards
- Homeworks
- Midterm
- Project
- Grading
- Participation
- Absenses and Accommodation

Do I really need to know this ____?

- Libraries are high-tech places
- Librarians foster IT literacy
- You need to buy and maintain IT systems

Do I really need to know this ____?

- Libraries are high-tech places
- Librarians foster IT literacy
- You need to buy and maintain IT systems

Do I really need to know this ____?

- Libraries are high-tech places
- Librarians foster IT literacy
- You need to buy and maintain IT systems

Ask questions

- In class
- On Piazza
- Office hours

Do I really need to know this ____?

I just started a new job on Monday as a Librarian at the National Archives. I was hired to catalog a special project with the archives and never thought I'd need anything from 690 . . . And then on the first day we started talking about the database of objects. When we finally got access to the database, I realized it was entirely a SQL database and had to use everything you taught us last semester to help think of how to run the queries and ask the designer what it is capable of.

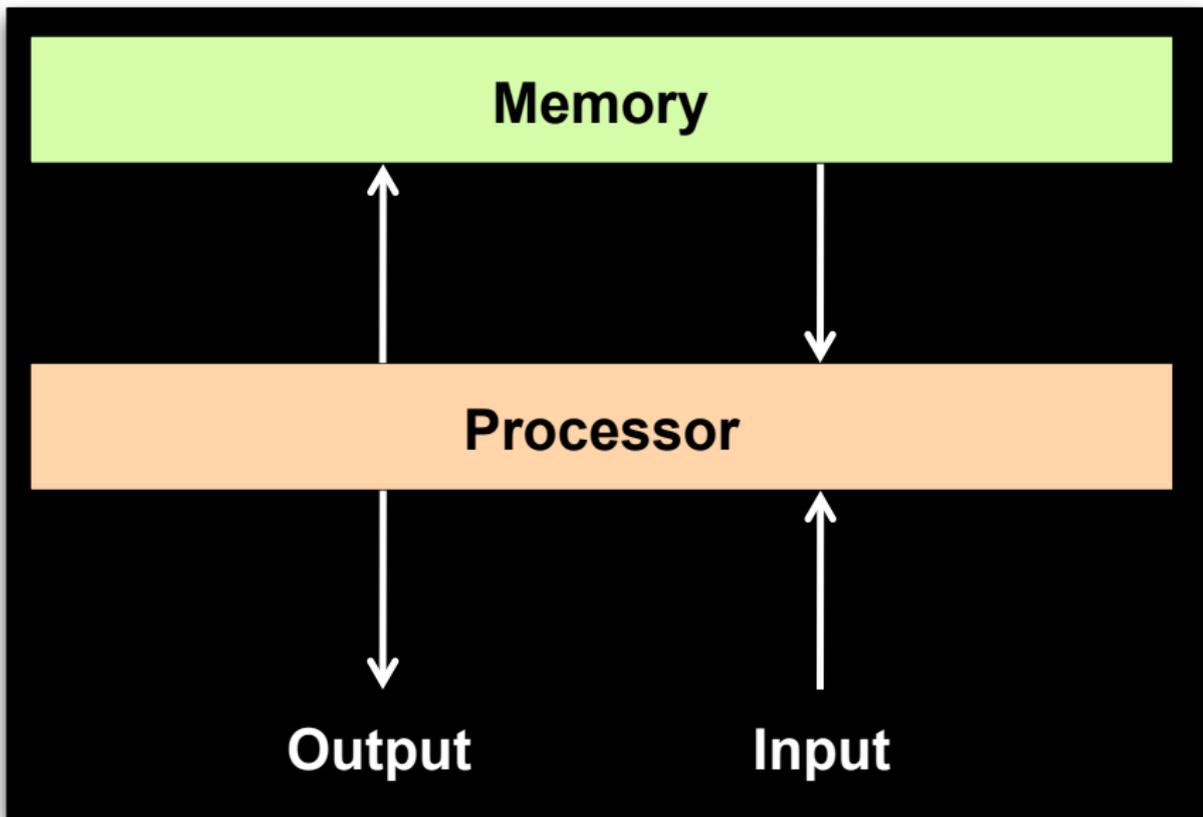
I just wanted to share because it's one of those times I never expected to realize how useful your class was.

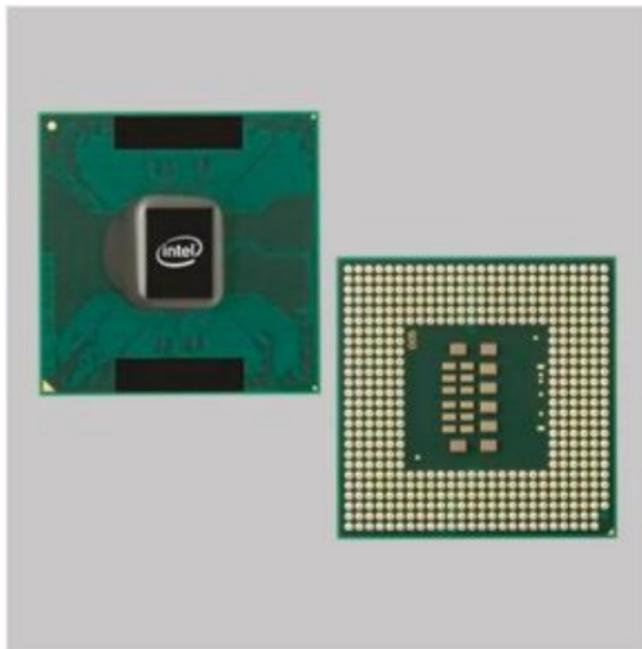
Do I really need to know this ____?

I actually had a job interview for a professional librarian job this week and it was about digital libraries. They asked me questions about what do I know about IP validation, HTML, servers, etc. Thankfully, fresh out of your class, I was able to say “actually, I know a TON about this stuff! There was a question on our midterm about IP validations and everything!”

Outline

- 1 History
- 2 Introductions
- 3 Computing**
- 4 Networking
- 5 How Big and How Fast?
- 6 How Fast?







Comparing Memory

- Quantity
- Speed
- Cost
- Volatility
- Physics

Outline

- 1 History
- 2 Introductions
- 3 Computing
- 4 Networking**
- 5 How Big and How Fast?
- 6 How Fast?

Why do you want to link computers together?

- Sharing data
- Sharing hardware
- Sharing software
- Increasing robustness
- Facilitating communications
- Facilitating commerce

How did the Internet start?

- ARPANET (Advanced Research Projects Agency)
- Operational in 1969: Sharing data between four universities (UCLA, UCSB, Stanford, Utah)
- Refers to the infrastructure that transfers data (c.f. “Web”)
- Next week . . . how it works now

Break!

Outline

- 1 History
- 2 Introductions
- 3 Computing
- 4 Networking
- 5 How Big and How Fast?**
- 6 How Fast?

Trends in Computing

- Faster
- Cheaper
- Smaller

Math diversion / refresher

- Division

$$\frac{2 \cdot 4}{4} = 2 \quad (1)$$

- Exponents

$$2^3 = 2 \cdot 2 \cdot 2 = 8 \quad (2)$$

- Scientific notation

$$526.6 \cdot 10^3 = 526.6 \cdot 1000 = 526600 \quad (3)$$

Math diversion / refresher

- Division

$$\frac{2 \cdot 4}{4} = 2 \quad (1)$$

- Exponents

$$2^3 = 2 \cdot 2 \cdot 2 = 8 \quad (2)$$

- Scientific notation

$$526.6 \cdot 10^3 = 526.6 \cdot 1000 = 526600 \quad (3)$$

Math diversion / refresher

- Division

$$\frac{2 \cdot 4}{4} = 2 \quad (1)$$

- Exponents

$$2^3 = 2 \cdot 2 \cdot 2 = 8 \quad (2)$$

- Scientific notation

$$526.6 \cdot 10^3 = 526.6 \cdot 1000 = 526600 \quad (3)$$

Bits, bytes, and why it's important

- You'll likely have to make decisions about computing resources
- Knowing how data converts into megabytes will help you know your needs
- Binary numbers appear in the real world (right)
- At the bottom, everything in the computer is binary



Counting in Binary

32	16	08	04	02	01	10	01
					1		1
				1	0		2
				1	1		3
		1	0	0			4
		1	0	1			5
		1	1	0			6
		1	1	1			7
	1	0	0	0			8
	1	0	0	1			9
	1	0	1	0		1	0
	1	0	1	1		1	1
	1	1	0	0		1	2
	1	1	0	1		1	3
	1	1	1	0		1	4
	1	1	1	1		1	5
1	0	0	0	0		1	6

How to convert from a binary number

- Write down powers of 2

1 2 4 8 16 32 64 128 256 512 1024 2048 4096 8192 ...

How to convert from a binary number

- Write down powers of 2

1 2 4 8 16 32 64 128 256 512 1024 2048 4096 8192 ...

- Take a number: 42
- Subtract out the largest power of 2 you can: 32
- You now know your binary number will look like 1XXXXX
- Repeat the process for each remaining bit left to right
 - ▶ If you can subtract, it's 1
 - ▶ If you can't subtract, it's 0

Power of 2	32	16	8	4	2	1
Remainder	10	10	2	2	0	0
bit	1	0	1	0	1	0

Everything on Your Computer is a Binary Number

01000001 = A

01000010 = B

01000011 = C

01000100 = D

01000101 = E

01000110 = F

01000111 = G

01001000 = H

01001001 = I

01001010 = J

01001011 = K

...

01100001 = a

01100010 = b

01100011 = c

01100100 = d

01100101 = e

01100110 = f

01100111 = g

01101000 = h

01101001 = i

01101010 = j

01101011 = k

...

Everything on Your Computer is a Binary Number

An HTML File

```
<HTML>  
</HTML>
```

How it's stored

Everything on Your Computer is a Binary Number

An HTML File

```
<HTML>  
</HTML>
```

How it's stored

```
00111100:
```

Everything on Your Computer is a Binary Number

An HTML File

```
<HTML>  
</HTML>
```

How it's stored

```
00111100:01001000:
```

Everything on Your Computer is a Binary Number

An HTML File

```
<HTML>  
</HTML>
```

How it's stored

```
00111100:01001000:01010100:
```

Everything on Your Computer is a Binary Number

An HTML File

```
<HTML>  
</HTML>
```

How it's stored

```
00111100:01001000:01010100:01001101:
```

Everything on Your Computer is a Binary Number

An HTML File

```
<HTML>  
</HTML>
```

How it's stored

```
00111100:01001000:01010100:01001101:  
01001100:
```

Everything on Your Computer is a Binary Number

An HTML File

```
<HTML>  
</HTML>
```

How it's stored

```
00111100:01001000:01010100:01001101:  
01001100:00111110:
```

Everything on Your Computer is a Binary Number

An HTML File

```
<HTML>  
</HTML>
```

How it's stored

```
00111100:01001000:01010100:01001101:  
01001100:00111110:00001010:00111100:  
00101111:01001000:01010100:01001101:  
01001100:00111110:00001010:00000100
```

Everything on Your Computer is a Binary Number

An HTML File

```
<HTML>  
</HTML>
```

How it's stored

```
00111100:01001000:01010100:01001101:  
01001100:00111110:00001010:00111100:  
00101111:01001000:01010100:01001101:  
01001100:00111110:00001010:0000100
```

Amounts of Data

Unit	Abbreviation	Size (bytes)
bit	b	1/8
byte	B	1
kilobyte	KB	

Amounts of Data

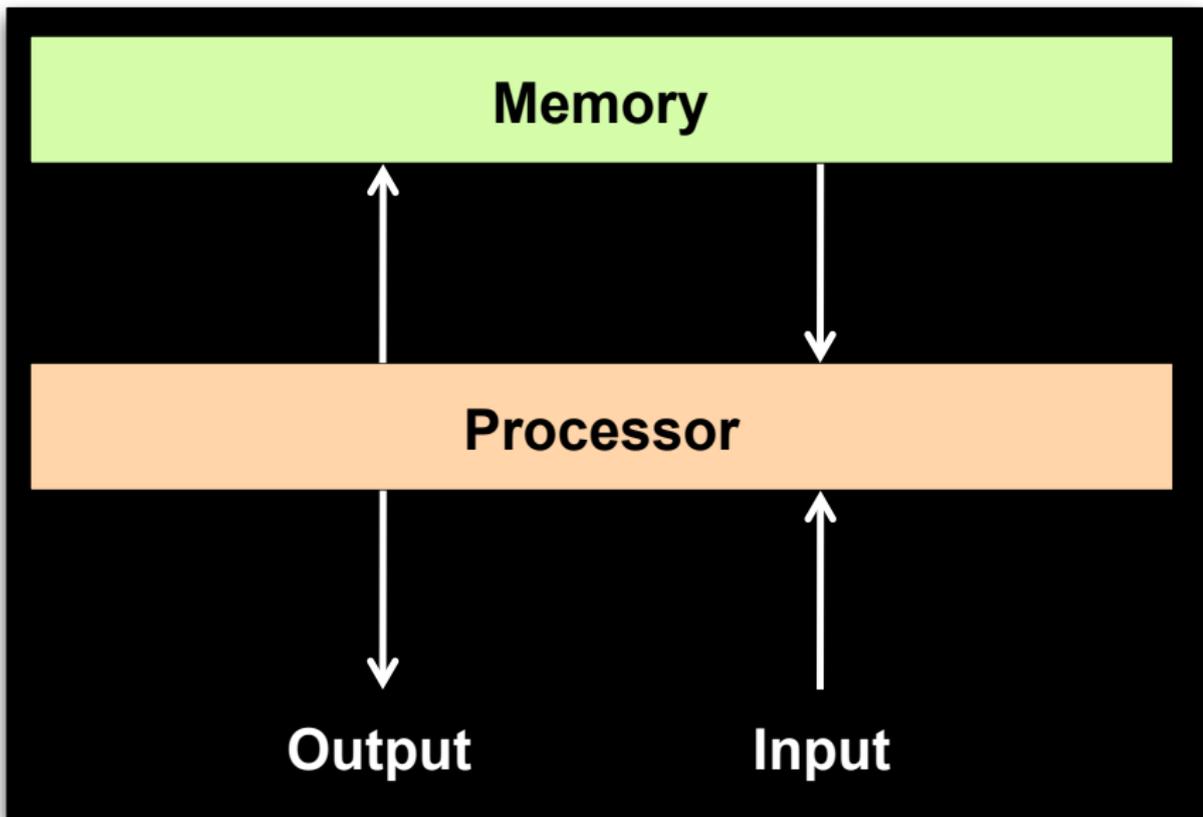
Unit	Abbreviation	Size (bytes)
bit	b	$1/8$
byte	B	1
kilobyte	KB	$2^{10} = 1,024$
megabyte	MB	$2^{20} = 1,048,576$
gigabyte	GB	$2^{30} = 1,073,741,824$
terabyte	TB	$2^{40} = 1,099,511,627,776$
petabyte	PB	$2^{50} = 1,125,899,906,842,624$

Outline

- 1 History
- 2 Introductions
- 3 Computing
- 4 Networking
- 5 How Big and How Fast?
- 6 How Fast?**

Moore's Law

The number of components on an integrated circuit will double every 18 months (1965)



Thinking about Speed

- Speed can be expressed in two ways:
 - ▶ How many things can you do in one second?
 - ▶ How long to do something once?
- Convenient units are typically used
 - ▶ 1 GHz instead of 1,000,000,000 Hz
 - ▶ 10 microseconds rather than 0.00001 seconds
- When comparing measurements, convert units first!

Clock Speeds

Unit	Abbreviation	Cycles per second
hertz	Hz	1
kilohertz	KHz	$10^3 = 1,000$
megahertz	MHz	$10^6 = 1,000,000$
gigahertz	GHz	$10^9 = 1,000,000,000$

Who's Faster

- Intel Pentium 2 (2000): 500 MHz
- Intel Celeron (2002): 2.10 GHz
- Intel Pentium 4 (2004): 3.80 GHz
- Intel Core 2 Duo (2008): 2.6 GHz
- Intel Xeon 8 Core (2010): 2.26 GHz

Units of Time

Unit	Abbreviation	Duration (seconds)
second	s	1
millisecond	ms	$10^{-3} = 1/1000$ s
microsecond	μ s	$10^{-6} = 1/1,000,000$ s
nanosecond	ns	$10^{-9} = 1/1,000,000,000$ s

Units of Time

Unit	Abbreviation	Duration (seconds)
second	s	1
millisecond	ms	$10^{-3} = 1/1000$ s
microsecond	μ s	$10^{-6} = 1/1,000,000$ s
nanosecond	ns	$10^{-9} = 1/1,000,000,000$ s

Light travels $0.3048m$ in a nanosecond.

Example

How long does it take a 1GHZ PC to sum a million numbers?

$$\frac{10^6 \text{ operations}}{10^9 \frac{\text{operations}}{s}} = 10^{-3} s = 1 \text{ ms} \quad (4)$$

In general ...

$$\frac{A^B}{A^C} = A^{B-C} \quad (5)$$



Typical Access Time: 50ns

Data moves at different speeds

- RAM: Fast and expensive
- HDD: Slow and cheap
- Corollary: People associate the sound of their hard drive with the computer being slow
- How to get the best of both worlds (hint: think of library and your bookshelf)

Caching

- Idea: move data you're going to use from slow memory into fast memory
- Slow memory is cheap so you can buy lots of it
- Caching gives you the illusion of having lots of fast memory
- How do we know what data to cache?
 - ▶ Spatial locality: If the system fetched x , it is likely to fetch data located near x (Why?)
 - ▶ Temporal locality: If the system fetched x , it is likely to fetch x again (Why?)

Big Picture

- Two parts of moving data from here to there:
 - ▶ Getting the first bit there (latency)
 - ▶ Getting everything there (bandwidth)
- Fundamentally, there's no difference:
 - ▶ Moving data from the processor to RAM
 - ▶ Saving a file to disk
 - ▶ Downloading pirated music from a server in China

Latency or bandwidth?

- Audio chat
- Downloading an mp3
- Streaming video (e.g., NPR broadcast over Web)

Recap

- History of computing
- Computers and networks
- Concepts of Space (how big?) and Time (how fast?)

Recap

- History of computing
- Computers and networks
- Concepts of Space (how big?) and Time (how fast?)

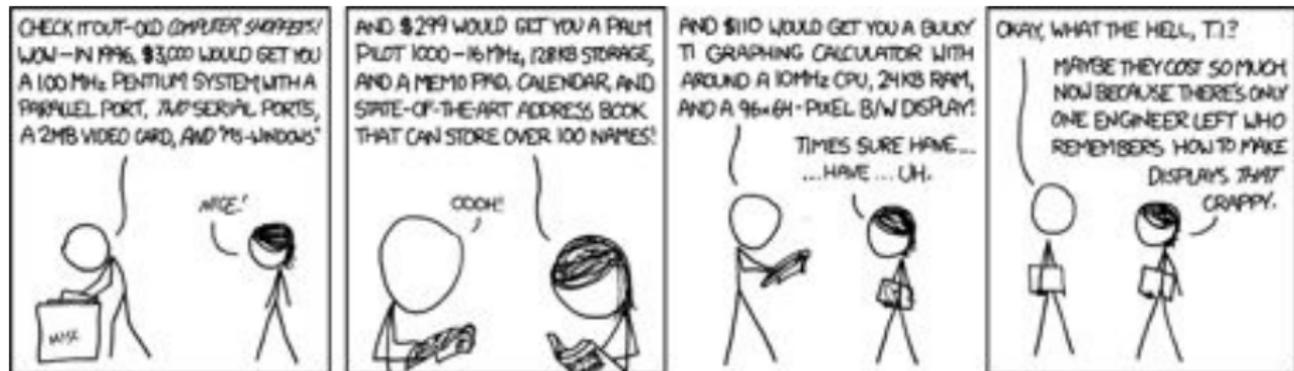
Next time ...

More nuts and bolts of the Web.

Assignment 0

- Not worth points
- Don't stress about it - if you have any problems, we'll tackle them in class next week
- Why Chrome?
 - ▶ Obeys standards
 - ▶ In past years, seemed to be the least problematic
 - ▶ Stays current
 - ▶ Works on OS X, Windows, Linux

Discussion



What's going on here?

Discussion

Is a phone a computer?