Bits and Bytes

At the smallest scale in the computer, information is stored as bits and bytes. In this section, we'll look at how that works.

Bit

- Bit, like an atom, the smallest unit of storage
- A bit stores just a 0 or 1
- "In the computer it's all 0's and 1's" ... bits
- Anything with two separate states can store 1 bit
- -Nick's tennis racket example
- Chip uses areas of electric charge as 0/1 states
- Hard drive uses spots North/South magnetism 0/1 states
- A bit is too small to be much use
- Group 8 bits together to make 1 byte

Everything in a computer is 0's and 1's ... what does that mean? The **bit** stores just a 0 or 1 .. it's the smallest building block of storage.

Byte

- One byte = grouping of 8 bits
- e.g. 0 1 0 1 1 0 1 0
- One byte can store one letter, e.g. 'A' or 'x'

How Many Patterns With N Bits?

How many distinct patterns can be made with 1, 2, or 3 bits?

Number of bits	Distinct Patterns
1	0 1
2	00 01 10 11
3	000 001 010 011 100 101 110 111

- 3 bits vs. 2 bits
- Consider just the leftmost bit
- It can only be 0 or 1
- Lefmost bit is 0, then append 2-bit patterns
- Leftmost bit is 1, then append 2-bit patterns again
- Result ... 3-bits has twice as many patterns as 2-bits

Number of bits	Distinct Patterns
1	0 1
2	00 01 10 11
3	000 001 010 011 100 101 110 111

- In general: add 1 bit, double the number of patterns
- 1 bit - 2 patterns
- 2 bits - 4
- 3 bits - 8
- 4 bits - 16
- 5 bits - 32

- 6 bits - 64
- 7 bits - 128
- 8 bits - 256 - one byte
- Mathematically: n bits yields 2^n patterns (2 to the nth power)

One Byte - 256 Patterns

- 1 byte is group of 8 bits
- 8 bits can make 256 different patterns
- How to use the 256 patterns?
- How to store a number in a byte?
- Start with 0, go up, one pattern per number, until run out of patterns
- 0, 1, 2, 3, 4, 5, ... 254, 255
- One byte can hold a number between 0 and 255
- i.e. with 256 distinct patterns, we can store a number in the range 0..255
- Code: pixel.setRed(n) took a number 0..255. Why?
- The red/green/blue image numbers are each stored in **one byte**

Bytes

- "Byte" - unit of information storage
- A document, an image, a movie .. how many bytes?
- 1 byte is enough to hold 1 typed letter, e.g. 'b' or 'X'
- Later we'll look at storage in: RAM, hard drives, flash drives

- All measured in bytes, despite being very different hardware
- **Kilobyte**, KB, about 1 thousand bytes
- **Megabyte**, MB, about 1 million bytes
- **Gigabyte**, GB, about 1 billion bytes
- **Terabyte**, TB, about 1 trillion bytes (rare)

The space that data takes up in the computer is measured in by the "byte". One byte is big enough to hold a single typed letter, like 'a'. Here we'll look at storing data in RAM memory and in persistent storage like a hard drive. All of that storage space will be measured in bytes. We'll look at byte arithmetic in more detail later.

Bytes and Letters - ASCII Code

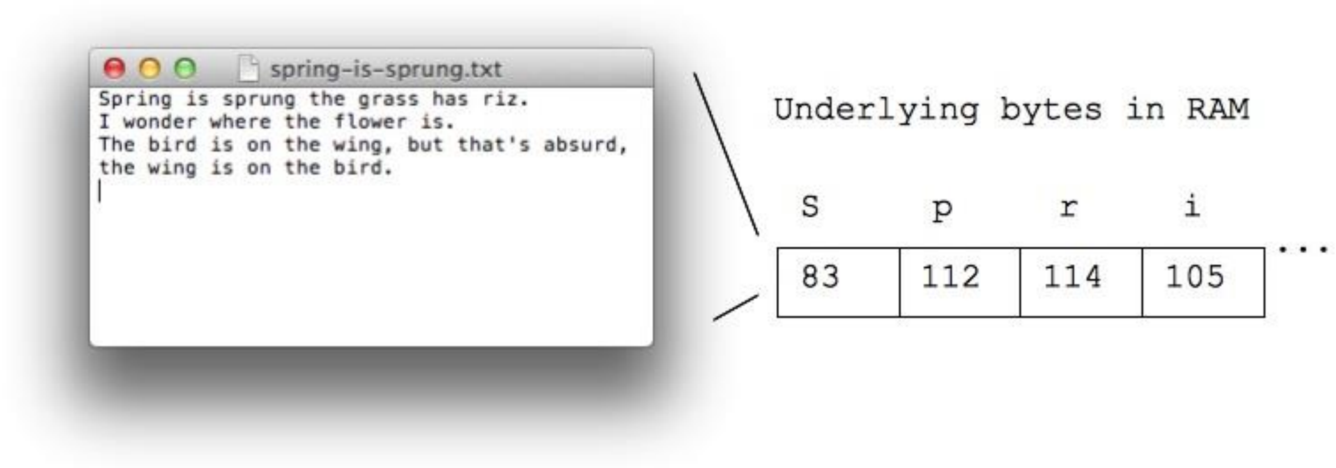
- ASCII is an encoding representing each typed letter by a number
- Each number is stored in one byte (so the number is in 0..255)
- A is 65
- B is 66
- a is 96
- space is 32
- "Unicode" is an encoding for mandarin, greek, arabic, etc. languages, typically 2-bytes per "letter"

```
32 space    47 /     62 >     77 M     92 \     107 k    122 z
33 !        48 0     63 ?     78 N     93 ]     108 l    123 {
34 "        49 1     64 @     79 O     94 ^     109 m    124 |
35 #        50 2     65 A     80 P     95 _     110 n    125 }
36 $        51 3     66 B     81 Q     96 `     111 o    126 ~
37%         52 4     67 C     82 R     97 a     112 p
38 &        53 5     68 D     83 S     98 b     113 q
39 '        54 6     69 E     84 T     99 c     114 r
40 (        55 7     70 F     85 U     100 d    115 s
41 )        56 8     71 G     86 V     101 e    116 t
42 *        57 9     72 H     87 W     102 f    117 u
43 +        58 :     73 I     88 X     103 g    118 v
```

44 ,	59 ;	74 J	89 Y	104 h	119 w
45 -	60 <	75 K	90 Z	105 i	120 x
46 .	61 =	76 L	91 [	106 j	121 y

Typing, Bytes, and You

- An example of bytes in your daily life
- When you type letters on your phone or computer
- Each letter is stored in a byte, as below
- 100 typed letters takes up 100 bytes
- When you send, say, a text message, the numbers are sent
- Text is quite compact, using few bytes, compared to images etc.



Kilobytes Megabytes Gigabytes

The size of information in the computer is measured in kilobytes, megabytes, gigabytes, and terabytes. In this section, we'll look at common sizes you would see in real life, and work some arithmetic. Any thinking person today should have a rough idea of what KB, MB and GB are.

Kilobyte or KB

- Kilobyte KB - about 1 thousand bytes

- As we know, 1 byte is one typed letter

  - see below for why the word "about" is required here
- A small email text is about 2 KB
- A 5 page paper might be 100 KB
- Text does not take a lot of bytes to store compared to images or video
- Math: if you have N bytes, that's N/1000 KB
- e.g. 23,000 bytes is about 23 KB

One kilobyte (KB) is a collection of about 1000 bytes. A page of ordinary roman alphabetic text takes about 2 kilobytes to store (about one byte per letter). A typical short email would also take up just 1 or 2 kilobytes. Text is one of the most naturally compact types of data at about one byte required to store each letter. In non-roman alphabets, such as Kanji, the storage takes up 2 or 4 bytes per "letter" which is still pretty compact compared to audio and images.

Megabyte or MB

- Megabyte (MB) - about 1 million bytes
- aka about 1000 KB
- MP3 audio is about 1 megabyte per minute
- A high quality digital picture is about 2-5 megabytes
- Math: if you have N KB, that's N/1000 MB
- e.g. 45,400 KB is 45.4 MB

One megabyte is about 1 million bytes (or about 1000 kilobytes). An MP3 audio file of a few minutes or a 10 million pixel image from a digital camera would typically take up few megabytes. The rule of thumb for MP3 audio is that 1 minute of audio takes up about 1 megabyte. Audio and image and video data typically stored in "compressed" form, MP3 being an example. We'll talk about how compression works later. A data CD disk stores about 700 MB. The audio on a CD is not compressed, which is why it takes so much more space than the MP3. The series of bits are represented as spiral path of tiny pits in the silver material in the disk. Imagine that each pit is interpreted as a 0, and the lack of a pit is a 1 as the spiral sequence is read. Fun fact: the whole spiral on a CD is over 5km long.

- Math - You Try It
- 2,000,000 bytes is about how many MB?
- 23,000 KB is about how many MB?
- 500 KB is about how many MB?

Solution

```
2 MB
23 MB
0.5 MB
```

Gigabyte or GB

- Gigabyte GB = about a billion bytes
- aka about 1000 MB
- GB is a common unit for modern hardware
- e.g. 4000 MB = 4 GB
- An ordinary computer might have:

  -4 GB or RAM

  -256 GB of persistent storage
- A DVD disk has a capacity 4.7GB (single layer)
- -Figure 2 GB per hour of video (varies greatly)
- A flash drive might hold 32 GB
- A small hard drive might hold 750 GB
- Math: if you have N MB, that's N/1000 GB

One gigabyte (GB) is about 1 billion bytes, or 1 thousand megabytes. A computer might have 4 GB of RAM. A flash memory card used in a camera might store 16 GB. A DVD movie is roughly 4-8 GB.

- Math - You Try It
- How many GB is 4,000,000,000 bytes?
- Say you have many 5 MB .jpeg images. How many fit on a 16 GB flash drive?

Solution

```
4 GB
5 MB per image means 200 images fit in 1 GB.
 200 * 16 is 12,000
```

Terabyte or TB

One terabyte (TB) is about 1000 gigabytes, or roughly 1 trillion bytes. You can buy 1 TB and 2 TB hard drives today, so we are just beginning the time when this term comes in to common use. Gigabyte used to be an exotic term too, until Moore's law made it common.

Gigahertz - Speed, not Bytes

One gigahertz is 1 billion cycles per second (a megahertz is a million cycles per second). Gigahertz is a measure of speed, very roughly the rate that at a CPU can do its simplest operation per second. Gigahertz does not precisely tell you how quickly a CPU gets work done, but it is roughly correlated. Higher gigahertz CPUs also tend to be more expensive to produce and they use more power (and as a result give off more heat) .. a challenge for putting fast CPUs in small devices like phones. The ARM company is famous for producing chips that are very productive with minimal power and heat. Almost all cell phones currently use ARM CPUs.

Kilobyte / Megabyte / Gigabyte Word Problems

You should be comfortable doing simple arithmetic to figure MB / GB sizes, just as you should be able to do basic computations with second, miles, kilos and so forth.

Basic plan: before adding measures X and Y, convert them to be in the same units.

Word Problems	**Solution**
Alice has 600 MB of data. Bob has 2000 MB of data. Will it all fit on Alice's 4 GB thumb drive?	Yes it fits: 600 MB + 2000 MB is 2600 MB. 2600 MB is 2.6 GB, so it will fit on the 4 GB drive no problem. Equivalently we could say that the 4 GB drive has

	space for 4000 MB.
Alice has 100 small images, each of which is 500 KB. How much space do they take up overall in MB?	100 times 500 KB is 50000 KB, which is 50 MB.
Your ghost hunting group is recording the sound inside a haunted Stanford classroom for 20 hours as MP3 audio files. About how much data will that be, expressed in GB?	MP3 audio takes up about 1 MB per minute. 20 hours, 60 minutes/hour, 20 * 60 yields 1200 minutes. So that's about 1200 MB, which is 1.2 GB.

Alternate Terminology: Kibibyte Mebibyte Gibibyte Tebibyte

It's convenient within the computer to organize things in groups of powers of 2. For example, 2^{10} is 1024, and so a program might group 1024 items together, as a sort of "round" number of things within the computer. The term "kilobyte" above refers to this group size of 1024 things. However, people also group things by thousands -- 1 thousand or 1 million items.

There's this problem with the word "megabyte" .. does it mean 1024 * 1024 bytes, i.e. 2^{20}which is 1,048,576, or does it mean exactly 1 million, 1000 * 1000. It's just a 5% difference, but marketers tend to prefer the 1 million, interpretation, since it makes their hard drives etc. appear to hold a little bit more. Also, the difference grows larger and larger for the gigabyte and terabyte sizes. In an attempt to fix this, the terms "kibibyte" "mebibyte" "gibibyte" "tebibyte" have been introduced to specifically mean the 1024 based units (see wikipedia kibibyte article). These terms do not seem to have caught on very strongly thus far. If nothing else, remember that terms like "megabyte" have this little wiggle room in them between the 1024 and 1000 based meanings. We will never grade off for this distinction .. "about a million" will be our close-enough interpretation for "megabyte". The "error" at the megabyte level is about 5%. At the terabyte level the error is about 10%.