Bit Basics

Eric McCreath

Bit Basics

A bit (Binary digIT) is single unit of binary storage. A bit is normally group with other bits to form a larger groupings of information.

4 bits form a nibble.

8 bits form a byte (e.g. 10100001 or 0xA1 in hex)

Bytes are grouped to form a word. These would normally be 1, 2, 4, or 8 bytes. The size of a word is depends on the particular processor.

Main memory can be thought of as a large array of bytes. However, often data needs to be properly aligned for the processor to use it.

Kilobyte, Megabyte, Gigabyte

When communicating with others about amounts of memory one needs to be careful as there is a number of different standards. Which standard is used depends on the context.

Multiples of bytes V · T · E

SI decimal prefixes		Binary usage	IEC binary prefixes	
Name (Symbol)	**Value**		**Name (Symbol)**	**Value**
kilobyte (kB/KB)	10^3	2^{10}	kibibyte (KiB)	2^{10}
megabyte (MB)	10^6	2^{20}	mebibyte (MiB)	2^{20}
gigabyte (GB)	10^9	2^{30}	gibibyte (GiB)	2^{30}
terabyte (TB)	10^{12}	2^{40}	tebibyte (TiB)	2^{40}
petabyte (PB)	10^{15}	2^{50}	pebibyte (PiB)	2^{50}
exabyte (EB)	10^{18}	2^{60}	exbibyte (EiB)	2^{60}
zettabyte (ZB)	10^{21}	2^{70}	zebibyte (ZiB)	2^{70}
yottabyte (YB)	10^{24}	2^{80}	yobibyte (YiB)	2^{80}

See also: Multiples of bits · Orders of magnitude of data

Storage on Modern Computers

My current desktop machine has:

- 6144 KiB cache
- 3.7 GiB main memory
- 11.0 GiB swap space reserved on the hard disk
- 283GiB of space on the hard disk
- CD-ROM stores around 703 MiB
- DVDs store from 4.3 to 8.0 GiM

Mbps, MIPS and FLOPS

Mbps means "mega bits per second" (1Mbps = 1000000 bits per second)

MIPS means "million instructions per second". Often a poor measure of processor performance, as this value is depends greatly on which instructions are executed.

FLOPS means "floating-point operations per second".

Characters

Characters can be stored using a number. This number can look up a standard table to determine the particular character. There are a few different standards.

ASCII (American Standar Code for Information Interchange) a 7 bit code which has a table of 128 characters. (see "man ascii").

EBCDIC (Extended Binary Coded Decimal Interchange Code) a 8 bit character encoding (used mainly on IBM mainframes).

UTF-8 is a variable width standard that can represent Unicode characters and is backward compatible with ASCII.

Strings

Strings are an artifact of the programming language.

Strings are normally stored as an array of characters with a null (or 0) terminating them.

What does the following mean?

0x43 0x4F 0x4D 0x50 0x32 0x33 0x30 0x30 0x00

Endianness

Endianness generally refers to the order the bytes within a single word are stored within main memory.

A big-endian machine will store most significate byte first (lowest address order). Processor include Motorala 6800, SPARC.

A little-endian machine will store the least significate byte first (lowest address order). Processors include x86, DEC Alpha, and Atmel AVR

Some hardware can be set so it can do either big-endian or little-endian on particular memory segments. These include ARM, PowerPC, MIPS, and IA-64.

Endianness

Generally a programmer will not care about this order as long as it is consistent. However, it becomes important when computers are connected via a network or share binary data.

Suppose we have a 32-bit machine and we stored the integer value for 10 at address 0xA0.

With a little-endian and big-endian machine we store the following:

little-endian	big-endian
$A0 : 0A$	$A0 : 00$
$A1 : 00$	$A1 : 00$
$A2 : 00$	$A2 : 00$
$A3 : 00$	$A3 : 0A$