

为什么要学编程基础

因为你首先是程序员，其次才是前端。

一个程序员需要知道

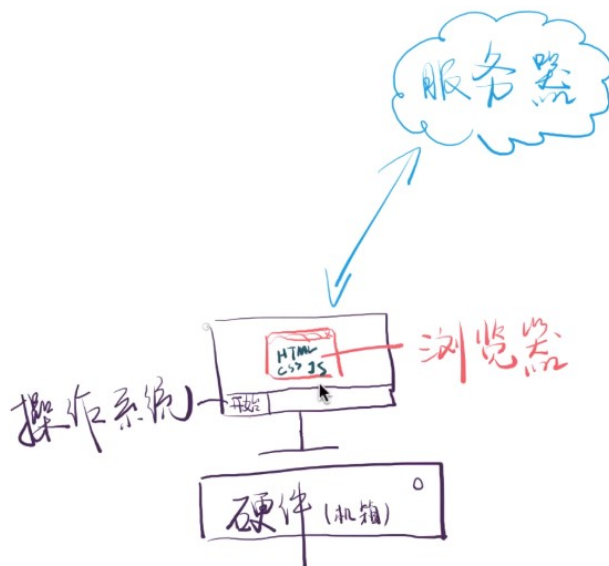
1. 硬件与软件：计算机的运行原理（《[编码](#)》）
2. 最大的软件：操作系统（[维基百科](#)）
3. 自己写软件：数据结构 & 算法（《[数据结构与算法分析](#)》）
4. 多人写软件：软件工程（[代码大全](#)）

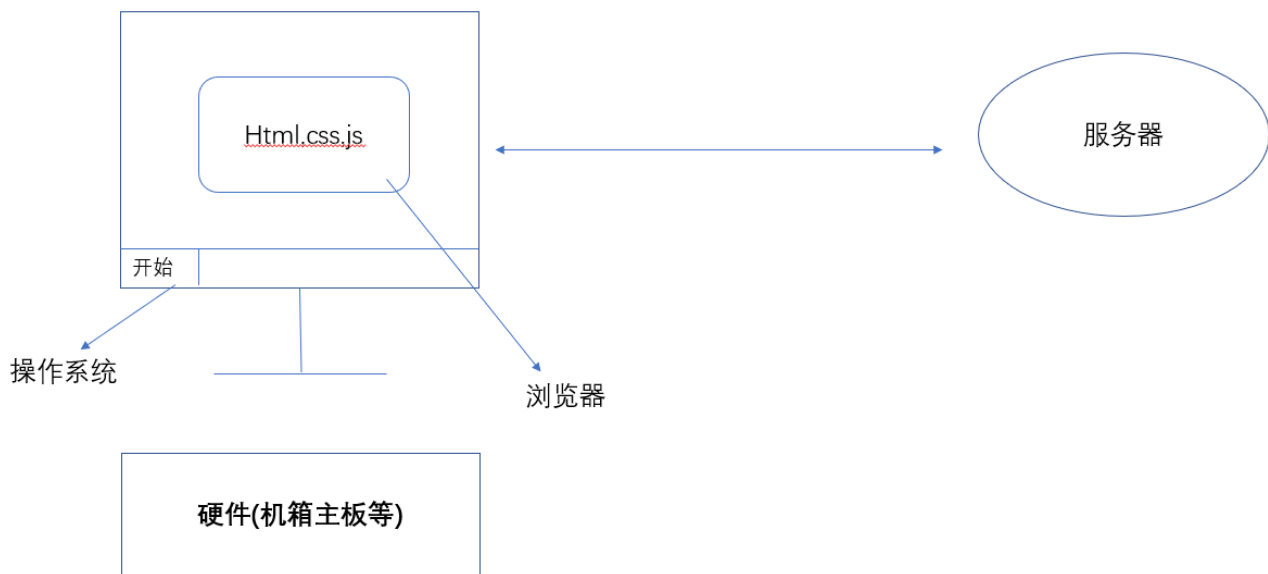
数据结构可以先把所有快排算法研究一波！

前端为什么学编程基础

- [操作系统](#) 运行于 [硬件](#) 之上
- [浏览器](#) 运行于 [操作系统](#) 之上
- HTML/CSS/JS 运行于 [浏览器](#) 之上
- HTML/CSS/JS 和 [数据](#) 都来自于 [服务器](#)

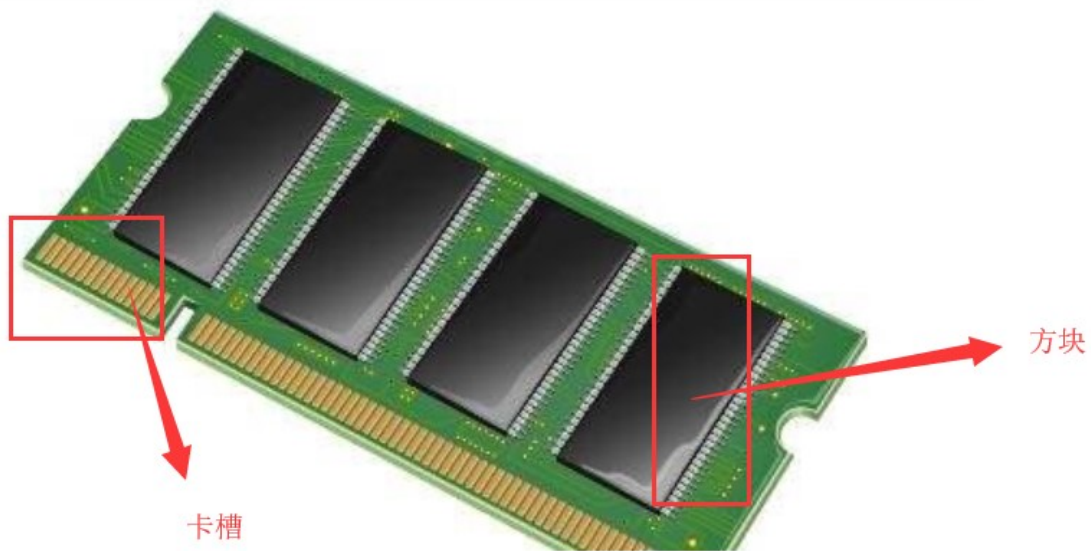
防止成为井底之蛙





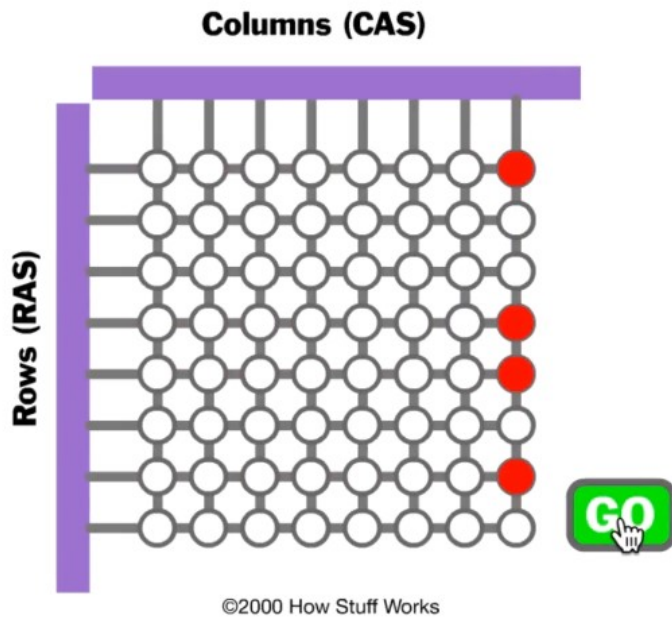
- 1.操作系统运行与硬件之上;
- 2.浏览器运行于操作系统之上;
- 3.html,css,js运行于浏览器之上;
- 4.html, css,js和数据皆来自于服务器;

内存条:



"方块"一般是4个,8个。。。;
一个方块里面有很多存储0和1的单位;(如下图:以8个为一排)

第一步：（内存）如何存储 0 和 1？

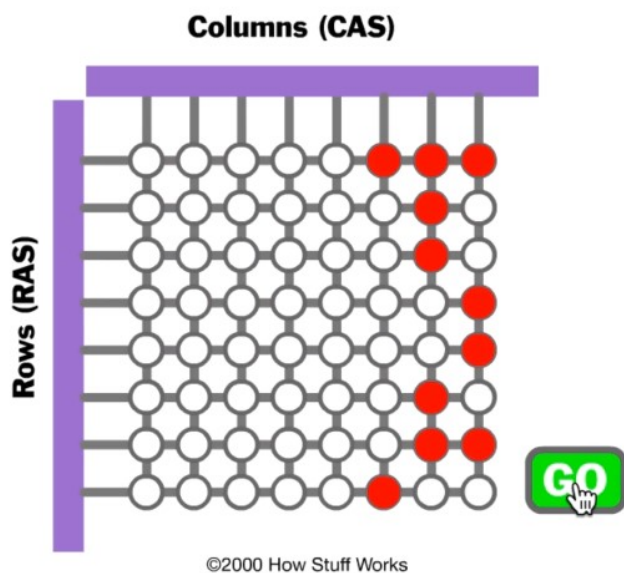


存储

- 1 就充电
- 0 就不充电

读取

- 电量大于 50% 就是 1
- 电量小于 50% 就是 0



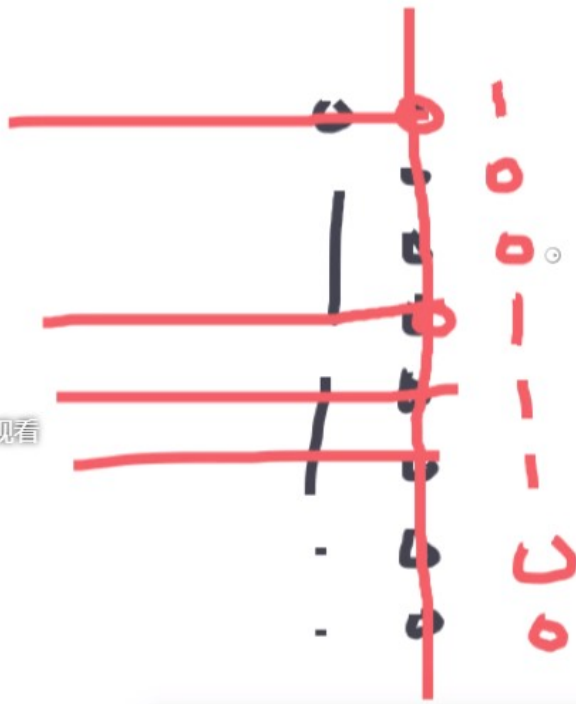
存储

- 1 就充电
- 0 就不充电

读取

- 电量大于 50% 就是 1
- 电量小于 50% 就是 0

75825正在观看



先选择列,再选择哪一行,选中的那一行设置为1图中有4行被选中,故有四个1;

如果我电脑的电池断电的话,马上就黑了,而后运行在内存里面的操作系统用也没了;all的0和1皆没了;为了解决耗电的问题,计算机要刷新,上图里面的红点比如它是1,为了使得其仍为1,那么就要在其消耗完它的电之前给其充电!

我们平常所说的电脑的cpu频率比如:2.1GHZ,就是说它每秒能充多少次电!每次充电时间很短,是纳秒;10的-9次方;

第二步：如何存储数字

十进制变二进制

$$37_{10} == 100101_2$$

$$-37_{10} == -100101_2$$

$$0.75_{10} == 0.11_2$$

计算机只存 0 和 1

为了方便书写,一般会将二进制数写为十六进制数

知识扩充:

负数会以补码的形式存储

小数会以浮点数的形式存储

请自行了解

第三步：如何存储字符

ASCII值	控制字符	ASCII值	控制字符	ASCII值	控制字符	ASCII值	控制字符
0	NUL	32	(space)	64	@	96	`
1	SOH	33	!	65	A	97	a
2	STX	34	"	66	B	98	b
3	ETX	35	#	67	C	99	c
4	EOT	36	\$	68	D	100	d
5	ENQ	37	%	69	E	101	e
6	ACK	38	&	70	F	102	f
7	BEL	39	,	71	G	103	g
8	BS	40	(72	H	104	h
9	HT	41)	73	I	105	i
10	LF	42	*	74	J	106	j
11	VT	43	+	75	K	107	k
12	FF	44	,	76	L	108	l
13	CR	45	-	77	M	109	m
14	SO	46	.	78	N	110	n
15	SI	47	/	79	O	111	o
16	DLE	48	0	80	P	112	p
17	DC1	49	1	81	Q	113	q
18	DC2	50	2	82	R	114	r
19	DC3	51	3	83	X	115	s
20	DC4	52	4	84	T	116	t
21	NAK	53	5	85	U	117	u
22	SYN	54	6	86	V	118	v
23	TB	55	7	87	W	119	w
24	CAN	56	8	88	X	120	x
25	EM	57	9	89	Y	121	y
26	SUB	58	:	90	Z	122	z
27	ESC	59	;	91	[123	{
28	FS	60	<	92	\	124	
29	GS	61	=	93]	125	}
30	RS	62	>	94	^	126	~
31	US	63	?	95	_	127	DEL

将每个字符编号

ASCII 美国信息交换标准代码

如果你想储存 a，那么就储存 97₁₀ 对应的二进制

a -> 0110 0001₂ -> 61₁₆

如果你想储存 1，那么就储存 49₁₀ 对应的二进制

1 -> 0011 0001₂ -> 31₁₆

扩大至16位:

第四步：如何存储中文

GB 2312 中国国家标准简体中文字符集

code	+0	+1	+2	+3	+4	+5	+6	+7	+8	+9	+A	+B	+C	+D	+E	+F
B6A0	丁	叮	叮	钉	顶	鼎	錠	定	订	丢	东	董	懂	动		
B6B0	栋	侗	冻	洞	兜	抖	斗	陡	豆	逗	痘	都	督	毒	牍	
B6C0	独	读	堵	睹	赌	杜	镀	肚	度	渡	妒	端	短	锻	段	断
B6D0	缀	堆	兑	队	对	墩	吨	敦	顿	囤	钝	盾	遁	援	哆	
B6E0	多	夺	垛	躲	朵	舵	刳	惰	堕	蛾	峨	鹅	俄	额	讹	
B6F0	娥	恶	厄	扼	遏	鄂	饿	恩	而	儿	耳	尔	饵	洱	二	
B7A0	贰	发	罚	筏	伐	乏	阉	法	珐	藩	帆	番	翻	樊	矾	
B7B0	钒	繁	凡	烦	反	返	范	贩	犯	饭	泛	坊	芳	方	肪	房
B7C0	防	妨	仿	访	放	非	非	啡	飞	肥	匪	诽	吠	肺	废	
B7D0	沸	费	芬	酚	氛	分	纷	坟	焚	汾	粉	奋	份	忿	愤	
B7E0	羹	丰	封	枫	峰	锋	风	疯	烽	逢	冯	缝	讽	奉	凤	
B7F0	佛	否	夫	敷	肤	孵	扶	拂	辐	幅	氟	符	伏	俘	服	
B8A0	浮	涪	福	袱	弗	甫	抚	辅	俯	釜	斧	脯	腑	府	腐	
B8B0	赴	副	覆	赋	复	傅	付	阜	父	腹	负	富	讣	附	妇	缚
B8C0	咐	噏	嘎	该	改	概	钙	盖	溉	干	甘	杆	柑	竿	肝	赶
B8D0	感	秆	敢	赣	冈	刚	钢	缸	肛	纲	岗	港	杠	篙	皋	高
B8E0	膏	羔	糕	搞	搞	稿	告	哥	歌	搁	戈	鸽	咯	疙	割	革
B8F0	葛	格	蛤	隔	铬	个	各	给	根	跟	耕	更	庚	羹		

GB2312 共收录 **6763 个汉字**，同时收录了包括拉丁字母、希腊字母、日文平假名及片假名字母、俄语西里尔字母在内的 **682 个字符**

后来为了存储生僻字、繁体字、日语、朝鲜语等，微软推出了 [GBK 字符集](#)

128237个字符需扩至 2^{32} (32位)

第五步：如何存储所有字符

将全球字符编号

Unicode 字符集

包括中日韩文字、藏文、盲文、楔形文字、[颜文字](#) :-)、绘文字😂

2016年6月时，Unicode 总共有 128237 个字符

第六步：如何将 Unicode 存到计算机里

以一种高性价比的方式

低性价比

a -> 00000000 00000000 00000000 01100001₂ = 0061₁₆

你 -> 00000000 00000000 01001111 01100000₂ = 4F60₁₆

高性价比 UTF-8

a -> 01100001

你-> 11100100 10111101 10100000

第六步：如何将 Unicode 存到计算机里

以一种高性价比的方式

低性价比

a -> 00000000 00000000 00000000 01100001₂ = 0061₁₆

你 -> 00000000 00000000 01001111 01100000₂ = 4F60₁₆

高性价比 UTF-8

a -> 01100001

你-> 11100100 10111101 10100000

你 里面的 1110 :表示后面还跟了两个字节,一共要读3个字节才能构成一个字符;

a 里面没有这样的前缀,说明是单字节!

UTF-8

UTF-8 是一种编码方式，不是字符集

00000000 00000000 00000000 01111111₂ 即 0000007F₁₆ 以下
0XXXXXXX

00000000 00000000 00000111 11111111₂ 即 000007FF₁₆ 以下
110XXXXX 10XXXXXX

00000000 00000000 11111111 11111111₂ 即 0000FFFF₁₆ 以下
1110XXXX 10XXXXXX 10XXXXXX

00000000 00011111 11111111 11111111₂ 即 001FFFFF₁₆ 以下
11110XXX 10XXXXXX 10XXXXXX 10XXXXXX

[更多](#)

UTF-8

UTF-8 是一种编码方式，不是字符集

00000000 00000000 00000000 01111111₂ 即 0000007F₁₆ 以下
0XXXXXXX

00000000 00000000 00000111 11111111₂ 即 000007FF₁₆ 以下
110XXXXX 10XXXXXX

00000000 00000000 11111111 11111111₂ 即 0000FFFF₁₆ 以下
1110XXXX 10XXXXXX 10XXXXXX

00000000 00011111 11111111 11111111₂ 即 001FFFFF₁₆ 以下
11110XXX 10XXXXXX 10XXXXXX 10XXXXXX

[更多](#)

表示3个字节

表示4个字节

为了节省空间,计算机内部采用utf-8的形式存储!!

小结

如何存储0和1--->存数字--->存字符--->存中文字符--->存all字符--->怎样以性价比高的方式将字符存入计算机(通过utf-8这种变长存储,可能是1个字节,可能2个字节,可能3个字节,也可能是4个字节)

js的bug

现实问题

编码问题

JavaScript 使用了 Unicode 字符集，但是没有使用 UTF-8 编码

JavaScript 用了 UCS-2 编码！

因为 1995 年 UTF-16 还没被发明出来，JavaScript 也不想使用 UTF-32

[具体看这里](#)

后果

ES5 无法表示 \uFFFF 之后的字符（如 \u1D306），某些情况下会出 bug

[更多](#)

Elements Console Sources Network Timeline Profiles Application Security Audits

top ▾ ☐ Preserve log

> '\u1d306'

< "06"

> |

js只能识别出两个字节;