

Základy forenzních databází

Tereza Uhlíková

verze y25

Kdo jsem

Tereza Uhlíková

Ústav analytické chemie

skupina teoretické spektroskopie

místnost A277

<https://web.vscht.cz/~uhlikovt/>

tereza.uhlikova@vscht.cz



Co bylo minule

- Co je to databáze
- Proč, kdo, kdy, jak ...
- Něco málo z historie
- CAP problém

O čem budeme mluvit dnes

- Základy informatiky
- Software
 - informace
 - záznam informace
 - číselné soustavy
 - písmenné kódy
- Hardware
 - Alan Turing, John von Neumann a počítač
 - historie vývoje počítače & super počítač

Proč informatika

- 1 technické obory
- 2 přírodní vědy a matematika (vývoj algoritmů, náročné výpočty a simulace)
- 3 medicína (vizualizace dat – CT, virtuální realita, informační systémy v medicíně...)
- 4 humanitní obory, vzdělávání – vytěžování znalostí, slovníky, automatické překladače, komunikace v přirozeném jazyce, e-learning, m-learning (mobilní vzdělávání, počítačů...)
- 5 ekonomické disciplíny (pokročilé statistické metody, datové sklady, dolování dat, elektronický obchod a bankovníctví...)
- 6 ...

Co je informatika

vědní obor, který se zabývá získáváním, zpracováním a využitím informace
angl. „Computer Science“

dnes vs. kdysi

kdysi - problém informace získat a získat je včas

dnes - problém vyhodnotit kvalitu a relevanci informace -> zdroj informace

Jaké vlastnosti by měla mít informace, která ovlivní vaše rozhodování?

- Správnost (jsou uvedeny zdroje?)
- Odbornost autora (jsou informace o autorovi?)
- Objektivnost (je dán prostor více stranám?)
- Aktuálnost (jsou informace datovány?)
- Ucelenost (je zpráva uvedena v kontextu?)

Co je informace

- předává se při komunikaci
- opak entropie
- různá definice v různých oborech:
 - informační věda a knihovnictví - sdělení, komunikovatelný poznatek, který má význam pro příjemce nebo údaj usnadňující volbu mezi alternativními rozhodovacími možnostmi.
 - exaktní vědy - sdělení, které vyhovuje přísným kritériím logiky či příslušné vědy.
 - ekonomická - sdělení, jehož výsledkem může být zisk nebo ztráta
 - výpočetní technika - kvantitativní vyjádření obsahu zprávy
- sémantika -> význam dat v databázi

Záznam informace

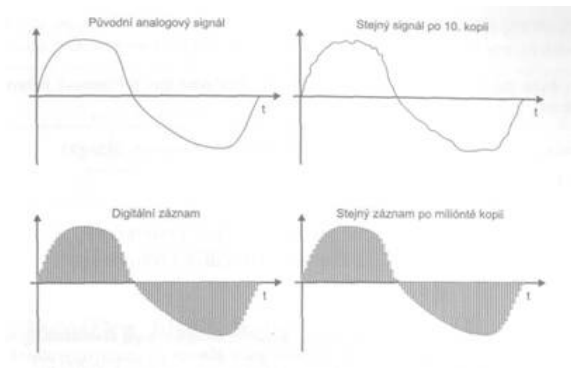


Záznam informace

strojový záznam

analogový - fyzikální veličina, dochází ke kompresi = ztrátě informace

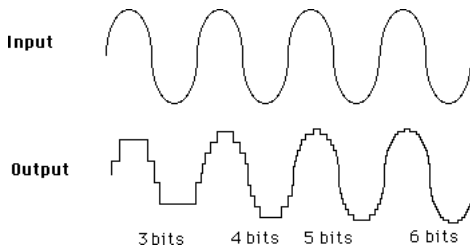
digitální - vzorkování a převod na binární číslo = minimální ztráta



Záznam a jednotka informace

jednotka

bit - rozhodnutí mezi dvěma alternativami (0, 1) binární hodnota



Byte - 8 bitů

? kolik různých hodnot může obsahovat 1B ?

Předpony

1 KiB (kilobajt) = 2^{10} B = 1 024 B

1 MiB(megabajt) = 2^{20} B = 1 048 576 B

1 GiB(gigabajt) = 2^{30} B = 1 073 741 824 B

1 TiB(terrabajt) = 2^{40} B = 1 099 511 627 780 B

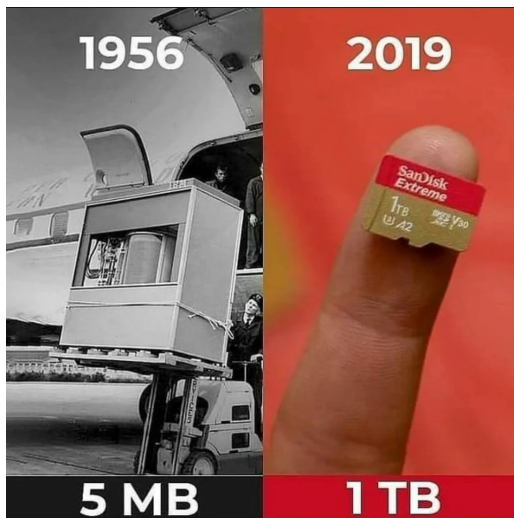
1 PiB(petabajt) = 2^{50} B = 1 125 899 907 000 000 B

1 EiB(exabajt) = 2^{60} B = 1 152 921 505 000 000 000 B

1 ZiB(zettabajt) = 2^{70} B = 1 180 591 621 000 000 000 000 B

1 YiB(yottabajt) = 2^{80} B = 1 208 925 819 614 629 174 706 176 B

Porovnání



<https://www.alza.cz/>

Číselné soustavy

číslo $\times$ číslice = abstraktní entita $\times$ znak pro zápis čísla
číselné soustavy

- poziční
- adiční (římská)

Dvojková (binární; obsahuje pouze číslice 0 a 1)

Osmičková (oktálová; číslice 0-7)

Šestnáctková (hexadecimální; 0-9, A-F)

Proč chceme dvojkovou soustavu?

Proč je záznam informace jako 0 a 1?

Převody

Číslo dělíme 2 a zbytek po dělení zapisujeme zprava.

Příklad:

$$2025 : 2 = 1012 \text{ zbytek } 1; 1012 : 2 = 506 \text{ zbytek } 0$$

$$506 : 2 = 253 \text{ zbytek } 0; 253 : 2 = 126 \text{ zbytek } 1$$

$$126 : 2 = 63 \text{ zbytek } 0; 63 : 2 = 31 \text{ zbytek } 1$$

$$31 : 2 = 15 \text{ zbytek } 1; 15 : 2 = 7 \text{ zbytek } 1$$

$$7 : 2 = 3 \text{ zbytek } 1; 3 : 2 = 1 \text{ zbytek } 1$$

$$2025_{(10)} = 1111101001_{(2)}$$

Dvojková soustava na desítkovou:

$$110: \text{ Zprava: } 1 \cdot 2^2 + 1 \cdot 2^1 + 0 \cdot 2^0 = 4 + 2 + 0 = 6_{(10)}$$

Převod hexadecimálního čísla na dekadické:

$$\text{Číslo } 7E9_{(16)} \text{ zapíšeme ve tvaru: } 7 \cdot 16^2 + E \cdot 16^1 + 9 \cdot 16^0$$

Dekadicky:

$$7 \cdot 16^2 + 14 \cdot 16^1 + 9 \cdot 16^0$$

$$1792 + 224 + 9 \cdot 1 = 2025$$

$$7E9_{(16)} = 2025_{(10)}$$

Bitová reprezentace reálných čísel

Desetinná čísla lze reprezentovat dvěma způsoby:

- v pevné řádové čárce (5 číslic před a 3 číslice za)
- v plovoucí řádové čárce:

Znaménko: Speciální bit (0 +, 1 -) určuje znaménko čísla

Mantisa: Představuje samotné číslice čísla

Exponent: Udává, o kolik míst se má desetinná čárka v mantise posunout, aby se získalo původní číslo.

→ Čím více bitů na mantisu, tím větší je přesnost uložení čísla.

→ Čím více bitů na exponent, tím větší je rozsah čísel, která lze uložit.

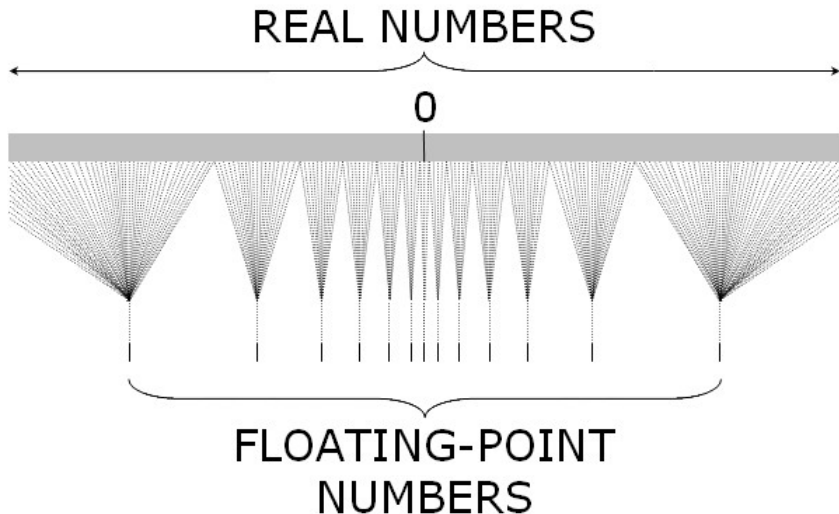
→ Kvůli omezenému počtu bitů je binární reprezentace desetinných čísel často nepřesná!

jednoduchá přesnost (single-precision) - 32 bitů

dvojitá přesnost (double-precision) - 64 bitů

⇒ win64 - Využívá 64 bitů k uložení čísla s plovoucí čárkou

Přesnost v desetinných číslech



Kódování

Morseova abeceda, Braillovo písmo, ...

zakódování X různých znaků = X různých dvojkových čísel

1 znak = 1 B

textová - znaková sada

ASCII (American Standard Code for Information Interchange) - 7bitový,

128 znaků, jeden znak jeden byte

UNICODE - UTF-8 - až 4B

ASCII

ASCII TABLE

Decimal	Hexadecimal	Binary	Octal	Char	Decimal	Hexadecimal	Binary	Octal	Char	Decimal	Hexadecimal	Binary	Octal	Char
0	0	0	0	[NULL]	48	30	110000	60	0	96	60	1100000	140	`
1	1	1	1	[START OF HEADING]	49	31	110001	61	1	97	61	1100001	141	a
2	2	10	2	[START OF TEXT]	50	32	110010	62	2	98	62	1100010	142	b
3	3	11	3	[END OF TEXT]	51	33	110011	63	3	99	63	1100011	143	c
4	4	100	4	[END OF TRANSMISSION]	52	34	110100	64	4	100	64	1100100	144	d
5	5	101	5	[ENQUIRY]	53	35	110101	65	5	101	65	1100101	145	e
6	6	110	6	[ACKNOWLEDGE]	54	36	110110	66	6	102	66	1100110	146	f
7	7	111	7	[BELL]	55	37	110111	67	7	103	67	1100111	147	g
8	8	1000	10	[BACKSPACE]	56	38	111000	70	8	104	68	1101000	150	h
9	9	1001	11	[HORIZONTAL TAB]	57	39	111001	71	9	105	69	1101001	151	i
10	A	1010	12	[LINE FEED]	58	3A	111010	72	:	106	6A	1101010	152	j
11	B	1011	13	[VERTICAL TAB]	59	3B	111011	73	;	107	6B	1101011	153	k
12	C	1100	14	[FORM FEED]	60	3C	111100	74	<	108	6C	1101100	154	l
13	D	1101	15	[CARRIAGE RETURN]	61	3D	111101	75	=	109	6D	1101101	155	m
14	E	1110	16	[SHIFT OUT]	62	3E	111110	76	>	110	6E	1101110	156	n
15	F	1111	17	[SHIFT IN]	63	3F	111111	77	?	111	6F	1101111	157	o
16	10	10000	20	[DATA LINK ESCAPE]	64	40	1000000	100	@	112	70	1110000	160	p
17	11	10001	21	[DEVICE CONTROL 1]	65	41	1000001	101	A	113	71	1110001	161	q
18	12	10010	22	[DEVICE CONTROL 2]	66	42	1000010	102	B	114	72	1110010	162	r
19	13	10011	23	[DEVICE CONTROL 3]	67	43	1000011	103	C	115	73	1110011	163	s
20	14	10100	24	[DEVICE CONTROL 4]	68	44	1000100	104	D	116	74	1110100	164	t
21	15	10101	25	[NEGATIVE ACKNOWLEDGE]	69	45	1000101	105	E	117	75	1110101	165	u
22	16	10110	26	[SYNCHRONOUS IDLE]	70	46	1000110	106	F	118	76	1110110	166	v
23	17	10111	27	[END OF TRANS. BLOCK]	71	47	1000111	107	G	119	77	1110111	167	w
24	18	11000	30	[CANCEL]	72	48	1001000	110	H	120	78	1111000	170	x
25	19	11001	31	[END OF MEDIUM]	73	49	1001001	111	I	121	79	1111001	171	y
26	1A	11010	32	[SUBSTITUTE]	74	4A	1001010	112	J	122	7A	1111010	172	z
27	1B	11011	33	[ESCAPE]	75	4B	1001011	113	K	123	7B	1111011	173	{
28	1C	11100	34	[FILE SEPARATOR]	76	4C	1001100	114	L	124	7C	1111100	174	
29	1D	11101	35	[GROUP SEPARATOR]	77	4D	1001101	115	M	125	7D	1111101	175	}
30	1E	11110	36	[RECORD SEPARATOR]	78	4E	1001110	116	N	126	7E	1111110	176	~
31	1F	11111	37	[UNIT SEPARATOR]	79	4F	1001111	117	O	127	7F	1111111	177	[DEL]
32	20	100000	40	[SPACE]	80	50	1010000	120	P					
33	21	100001	41	!	81	51	1010001	121	Q					
34	22	100010	42	"	82	52	1010010	122	R					
35	23	100011	43	#	83	53	1010011	123	S					
36	24	100100	44	\$	84	54	1010100	124	T					
37	25	100101	45	%	85	55	1010101	125	U					
38	26	100110	46	&	86	56	1010110	126	V					
39	27	100111	47	'	87	57	1010111	127	W					
40	28	101000	50	(	88	58	1011000	130	X					
41	29	101001	51	)	89	59	1011001	131	Y					
42	2A	101010	52	*	90	5A	1011010	132	Z					
43	2B	101011	53	+	91	5B	1011011	133	[					
44	2C	101100	54	,	92	5C	1011100	134	\					
45	2D	101101	55	-	93	5D	1011101	135	]					
46	2E	101110	56	.	94	5E	1011110	136	^					
47	2F	101111	57	/	95	5F	1011111	137	_					

Kódování obrazu a zvuku

kodování obrazu



- vektorová grafika - analýza geometrie

0	000	černá
1	001	modrá
2	010	zelená
3	100	červená
4	110	žlutá
5	101	magenta
6	011	cyan
7	111	bílá

- rastrový obrázek - pixel

kodování zvuku (vaw, mp3, flac)

Šifrování

základní (traspozice, substituce) - Ceasarova šifra

moderní - RSA (Rivest, Shamir, Adleman) - rozklad na prvočísla

- symetrické - stejný klíč
- asymetrické - dva jiné klíče - soukromý a veřejný

šifrovaná i zpráva

SSH - Secure Shell Protocol

Whatsapp - end-to-end - Curve25519 algoritmus - "Elliptic Curve Diffie Hellman algoritmus is a mathematical algorithm which helps two communicating entities to agree up on a shared secret without actually sending the actual keys to each other."

Komprese

Komprese - zmenšuje velikost souborů a složek, kompresní poměr
princip

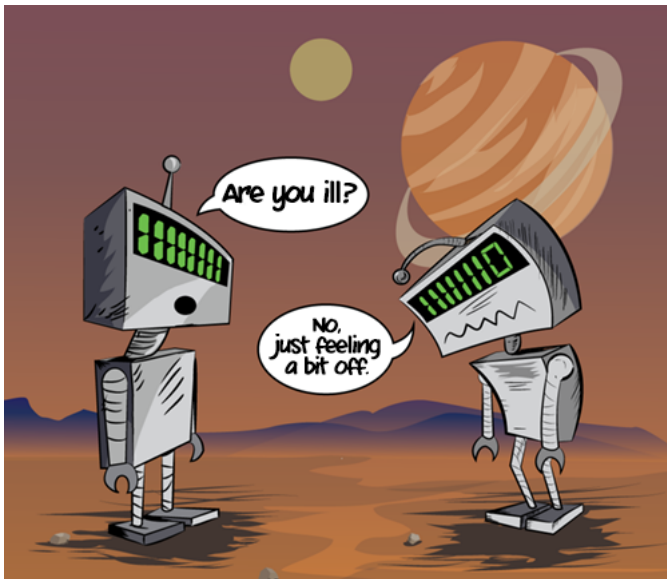
v počítači jsou všechny informace kódovány pomocí nul a jedniček – (např. 0010111010110001010110 atd.)

→ často se opakující úseky (např. čtveřice 0011), lze nahradit kratším úsekem

kompresní algoritmy - bzip2, gzip, zip, LZ4, xz, rar, 7z, LHa, ...

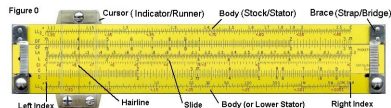
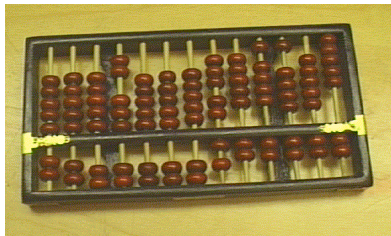
- ztrátová - některé informace jsou trvale ztraceny → psychovizuální nebo psychoakustický model. MP3, OGG, WMA, WMV, MPEG, DivX, JPEG, ...
- bezztrátová - zip, rar, tiff, png, gif, flac, wma, ...

a bit off



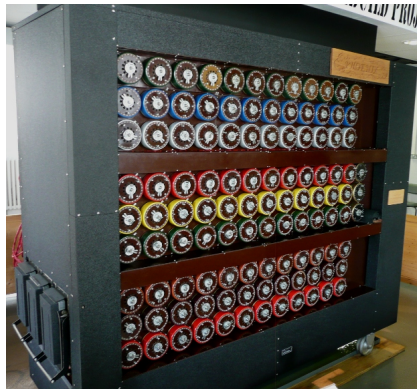
Historie vývoje počítače

- Abakus, logaritmické pravítko
- 1833 Charles Babbage
- 1936 Turingův stroj
- 1945 von Neumannovo schéma
- 1965 Moorův zákon



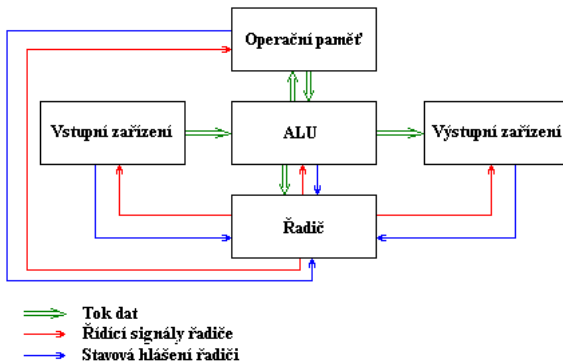
Alan Turing (1912-1954)

- definuje matematický model univerzálního stroje schopného vykonávat jakýkoliv popsateľný algoritmus
- definuje princip vypočitatelnosti
- koncept symbolického zpracování
- Turingův test

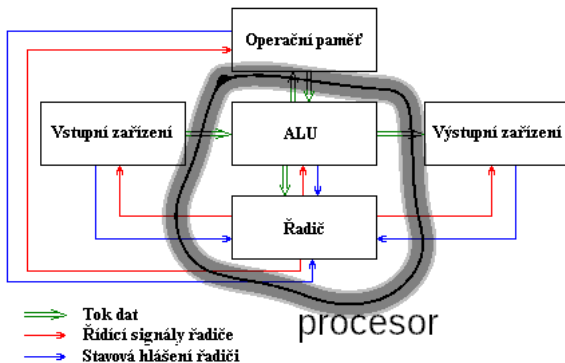


John von Neumann (1903-1957)

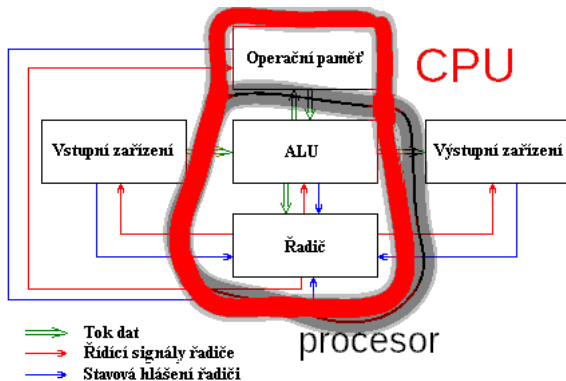
https://en.wikipedia.org/wiki/John_von_Neumann



Von Neumannovo schéma



Von Neumannovo schéma



Von Neumannovo schéma

- Operační paměť : slouží k uchování zpracovávaného programu, zpracovávaných dat a výsledků výpočtu
- ALU - Arithmetic-logic Unit (aritmetickologická jednotka): jednotka provádějící veškeré aritmetické výpočty a logické operace. Obsahuje sčítačky, násobičky (pro aritmetické výpočty) a komparátory (pro porovnávání)
- Řadič: řídicí jednotka, která řídí činnost všech částí počítače. Toto řízení je prováděno pomocí řídicích signálů, které jsou zasílány jednotlivým modulům. Reakce na řídicí signály, stavy jednotlivých modulů jsou naopak zasílány zpět řadiči pomocí stavových hlášení
- Vstupní zařízení: zařízení určená pro vstup programu a dat.
- Výstupní zařízení: zařízení určená pro výstup výsledků, které program zpracoval
- Procesor: Řadič + ALU
- CPU - Central Processor Unit (centrální procesorová jednotka):
Procesor + Operační paměť

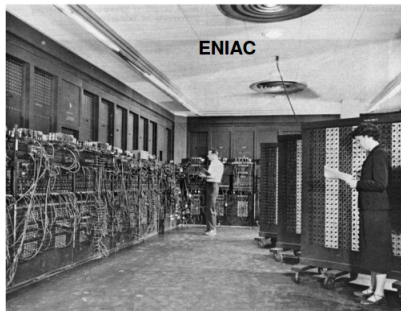
První generace 1944

elektronkové počítače, strojový kód, assembler autokód, děrné štítky, desítky skříní, operační rychlost op./s 10^2 - 10^4 , vnitřní paměť 1-2 KB

Elektronka



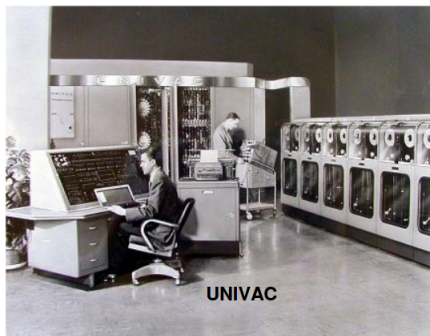
Kondenzátory



Druhá generace 1951

tranzistory, vyšší programovací jazyky FORTRAN, ALGOL, COBOL, do deseti skříní, operační rychlost op./s $10^4 - 10^5$, vnitřní paměť 16-32 KB

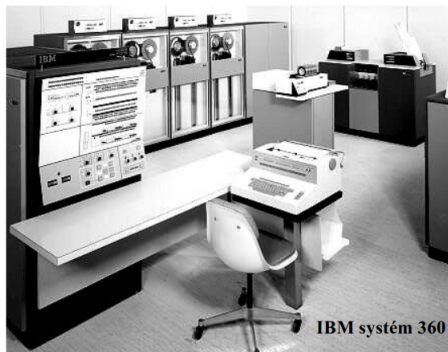
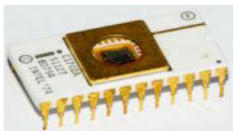
Tranzistory



Třetí generace 1965

integrováný obvod, operační systémy a modernější programovací jazyky PASCAL, LISP, do pěti skříní, operační rychlost op./s 10^5 - 10^6 , vnitřní paměť 0,5-2 MB dále až 16MB

Integrovaný obvod



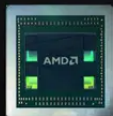
Čtvrtá generace

mikroprocesor, operační systémy a jazyky přizpůsobené uživatelům, jedna skříň, operační rychlost op./s 10^5 - 10^6 , vnitřní paměť 1-18 MB



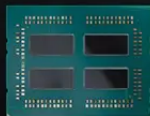
Nyní 3D technologie

AMD LEADERSHIP PACKAGING



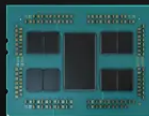
2.5D HBM

2015



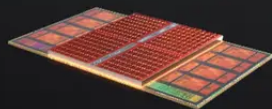
**MULTICHIP
MODULE**

2017



CHIPLETS

2019



3D CHIPLETS
(Chiplet + Advanced 3D Stacking)

2021

Led Industry in HBM, 2.5D & Chiplet Architecture | Aggressive Roadmap for Chiplet & 3D Integration

Rok 2025

www.top500.org

Rank	Site	System	Cores	Rmax (TFlop/s)	Rpeak (TFlop/s)	Power (kW)
1	DOE/NNSA/LLNL United States	El Capitan - HPE Cray EX255a, AMD 4th Gen EPYC 24C 1.8GHz, AMD Instinct MI300A, Slingshot-11, TOSS HPE	11,039,616	1,742.00	2,746.38	29,581
2	DOE/SC/Oak Ridge National Laboratory United States	Frontier - HPE Cray EX235a, AMD Optimized 3rd Generation EPYC 64C 2GHz, AMD Instinct MI250X, Slingshot-11, HPE Cray OS HPE	9,066,176	1,353.00	2,055.72	24,607
3	DOE/SC/Argonne National Laboratory United States	Aurora - HPE Cray EX - Intel Exascale Compute Blade, Xeon CPU Max 9470 52C 2.4GHz, Intel Data Center GPU Max, Slingshot-11 Intel	9,264,128	1,012.00	1,980.01	38,698
4	EuroHPC/FZJ Germany	JUPITER Booster - BullSequana XH3000, GH Superchip 72C 3GHz, NVIDIA GH200 Superchip, Quad-Rail NVIDIA InfiniBand NDR200, RedHat Enterprise Linux EVIDEN	4,801,344	793.40	930.00	13,088
5	Microsoft Azure	Eagle - Microsoft NDv5, Xeon	2,073,600	561.20	846.84	



El Capitan



IT4 Ostrava

<https://www.it4i.cz/>

Rank	site	cores	Rmax(Tf/s)	Rpeak(Tf/s)	power(kW)
1	El Capitan	11,039,616	1,742,000.00	2,746,376.09	29,580.0
195	Karol1naGPU	71,424	6,752.00	9,080.18	310.99
421	Karol1naCPU	92,160	2,837.40	3,833.86	502.70
...	PC - Mira	24	~1	~1	0.250
	Mobile s25	8			

Hardware - stručné opakování

Hardware

procesor - Intel a AMD + chladič a větrák

základní deska - s instalovaným Basic Input Output Systémem (BIOS)

sběrnice, řadiče

zdroj

operační paměť

pevné disky - HDD a Solid state

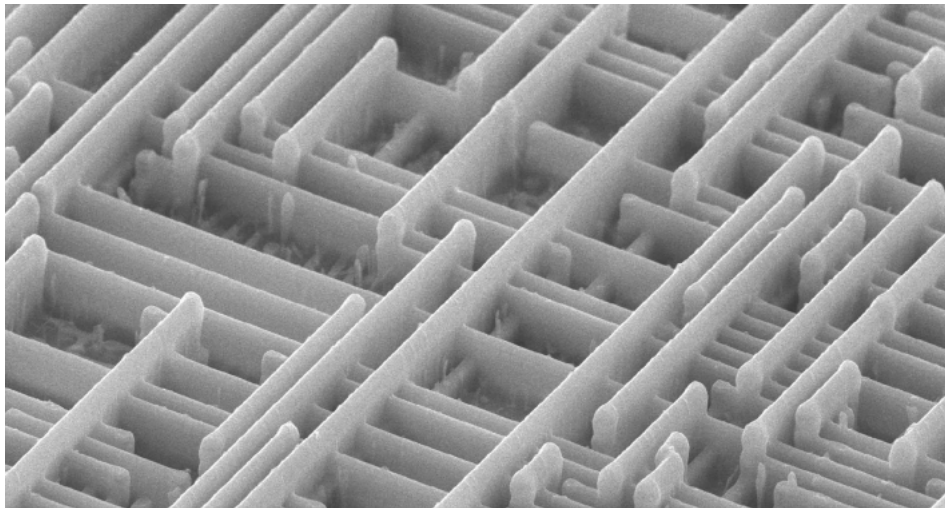
záznamová média

grafická karta

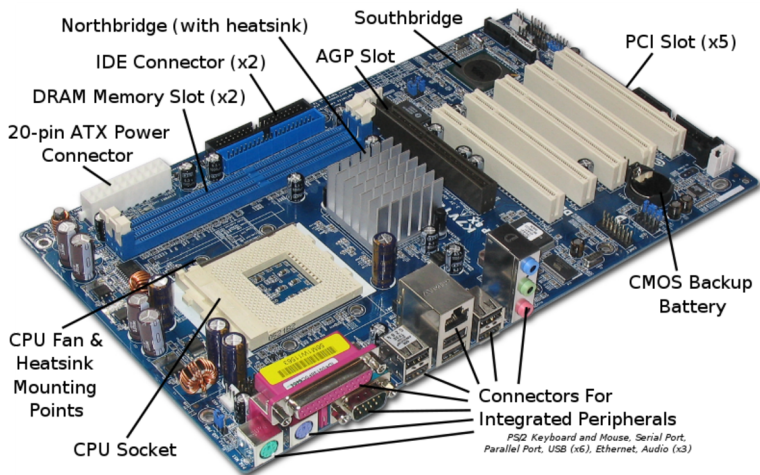
síťová karta

Processor

<https://www.youtube.com/watch?v=bor0qLifjz4>



motherboard



Úložná media

- neelektronické - papyrus, pergamen, papír, kámen, tabule, fotografická deska
- magnetické - disketa, pevný disk, videokazeta,...
- optické - CD, DVD, Blu-ray, ...
- elektronické - USB flash, SSD, paměťové karty, ...

1805 Děrný štítek -

Joseph-Maria Jacquard

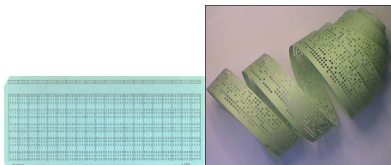
1878 Magnetická páska - Oberlin Smith

1967 8" disketa IBM 80 kB

1976 5.25" disketa

1984 3.5" disketa 1.44 MB

1995 ZIP 3.5" 100MB



1957 první tvrdý disk - IBM



HD



HD x SSD

hard disk se uvnitř otáčí

pevný disk



Skripta, stránky a materiály

http://projekty.fs.vsb.cz/463/edubase/VY_01_042/Z%C3%A1klady%20informatiky.pdf
<http://home.zcu.cz/~pkral/texty.pdf>
<https://www.konvoj.cz/Nakladatelstvi/Knihy/100289-haluza-rybicka-hala-scriptum-43-uvod-do-informatiky/100289-uvod-do-informatiky.pdf>
http://www.bigyzt.cz/shared/clanky/2893/ICT-Pripravy/IS4-Pripravy_informace.pdf
<http://www.ivt.mzf.cz/informatika-v-kostce/zakladni-pojmy/kodovani-a-sifrovani/>
<https://www.itnetwork.cz/hardware-pc/principy-pocitacu/reprezentace-cisel-v-pocitaci>
<https://www.napocitaci.cz/33/symetricke-a-asymetricke-sifrovani-uniqueidgOkE4NvrWuNY54vrLeM677jX7sp3Lu-ZpLpGVMy1prA/>
https://en.wikipedia.org/wiki/Charles_Babbage
<https://en.wikipedia.org/wiki/FLOPS>
<https://explodingtopics.com/blog/data-generated-per-day>
<https://www.google.com/about/datacenters/>

odkazy na obrázky

obrázky:

<https://comic.browserling.com/tag/hardware>

zaznam <https://www.gym-karvina.cz/ivt/okruhy/informatika.htm>

porovnání <https://www.zmescience.com/feature-post/technology-articles/computer-science/how-big-data-can-get/>

ascii <https://simple.m.wikipedia.org/wiki/File:ASCII-Table.svg>

<https://vektorova-grafika.cz/rastrova-a-vektorova-grafika-jaky-je-rozdil/>

<https://www.fi.muni.cz/usr/jkucera/pv109/xdavidov.html>

<https://commons.wikimedia.org/wiki/File:VNEUMAN.GIF>

<https://leonardo-supercomputer.cineca.eu/hpc-system/>

<https://www.datacenterknowledge.com/supercomputers/lawrence-livermore-national-lab-begins-building-el-capitan-supercomputer>