

ALIGNED RESEARCH GROUP

Working with data at a low level

How computer helps you to deal with data stuff

Nikolay Markov, 2020



Shameless plug

- My name is <u>Nikolay Markov</u> and my nick is <u>@enchantner</u> almost everywhere on the Internet
- I work as a Principal Architect at <u>Aligned Research</u> <u>Group</u>
- For more than 9 years I'm using Python, also Golang, C/C++, Rust and Scala
- Dealing with clouds, distributed computing and networking (AWS, Azure, Kubernetes)
- Writing <u>articles</u>, reading <u>lectures</u>, organizing <u>events</u>
- Like chatting about system design, architecture and other low-level magic



Small: never use for, iterrows(), itertuples() и apply()

for loop: 645 ms ± 31 ms per loop (mean ± std. dev. of 7 runs, 1 loop each)

iterrows(): 166 ms ± 2.42 ms per loop (mean ± std. dev. of 7 runs, 1 loop each)

itertuples(): <u>https://medium.com/@formigone/stop-using-df-iterrows-2fbc2931b60e</u>

apply(): 90.6 ms ± 7.55 ms per loop (mean ± std. dev. of 7 runs, 10 loops each)



SIMD & BLAS

What if we will apply one operation to multiple values?

<u>cuBLAS</u>

MMX/SSE/AVX...

NeCPP - BLAS discussion



Numpy + Intel MKL (+Install guide)

Multiplying matrices

OpenBLAS

pic source

Intel and its magic

Awesome article on AVX (in russian)

Skylake Overview – Power Management View

- PCH eDRAM x16 PCI DMI/OPI Express' PCle System Agent EC 2ch PECI DDR SVID PCU VR IMC Display Displa Core Core Core Core **Processor Graphics** 2/3/4 Note: Not to scale
 - Skylake is a SoC consisting of:
 - 2-4 CPU cores, Graphics, media, Ring interconnect, cache
 - Integrated System Agent (SA)
 - On package PCH and eDRAM
 - Improved performance with aggressive power savings
 - Package Control Unit (PCU) :
 - Power management logic and controller firmware
 - Continues tracking of internal statistics
 - Collects internal and external power telemetry: iMon, Psys
 - Interface to higher power management hierarchies: OS, BIOS, EC, graphics driver, DPTF, etc.

Intel[®] Architecture, Code Name Skylake



Compiling uncompilable

<u>Cython</u>



pic source

Linux BCC

Ways of compiling things

An interpreter

An Ahead-of-Time (AOT) compiler

An optimizing AOT compiler

:0:

0

A Just-in-Time (JIT) compiler

Gimme a normal compiler, please



<u>(pic source)</u>



Tracing JIT + PyPy Paper



LuaJIT (old, but gold)

Python world: PyPy (RPython), also Numba

100

News from Java - GraalVM (+explanation)



Interesting stuff to read - DMA

Using CPU to access memory is expensive

Imagine if we would have to do it for every network packet on a highly loaded server

Wouldn't it be easier to read/write network card directly?

(pic sourc

10

Use userspace TCP stack if you want to get more crazy

Interesting stuff to read - **SMP**

All processors are equal and homogenous

All CPUs have access to the same memory



Interesting stuff to read - NUMA

Every processor has its own memory

Great performance benefits if used right

All problems of a distibuted infrastructure (cache coherency, etc.) on one machine

pic source



Let multiple devices to have direct memory block exchange



Proprietary and works only from 2080+, but may bring substantial benefit for neural networks training & inference





Interesting stuff to read - USL



Before 1988 - CISC & page faults

The slowest operation is paging (aka swap)

Performance = # of page faults × # of instructions

The second second sequences of the second se

Bernstein (2001) 1. Organization and a decimination of the second second second second second additional second marker from lapshile date in constrained and the standard second se

RAM is really slow

CISC - Complex Instruction Set Computer - "let's put all the hard part inside the chip and just use it"

Data Locality is of utmost importance

A second process of process of the second process of the second

The same shall offer, here the lad parts and the shall parts and the shall be shall

Profile giver maxime chosepolitics inmaine i kulterna, to preme, kan posterna, poste income na nguesa cispate induction breakment regio trond. Chanaka. Cho-Sada ye na visite gabetak angu mayih nanga msitem hata poj e orazistan andre statekon tugih angura na hata poj e orazistan andre statekon tugih angura na posterna konstan terteka ata unida na presedenden state po profilerare i unida, anguranden angurande state por posterpar i unida, anguranden angurande statekon tertekon atanga posteria tugiha atanga statekon tertekon atanga posteria atanga statekon tertekon atanga posteria atanga statekon atanga posteria atanga posteria atanga posteria atanga statekon atanga posteria atanga posteria atanga posteria atanga statekon atanga posteria atanga posteria atanga posteria atanga posteria atanga statekon atanga posteria atanga posteria atanga posteria atanga statekon atanga posteria atanga posteria atanga posteria atanga posteria atanga statekon atanga posteria atan

vas uportili, vernite. Varna je soski minut dragosjen ni i belje ki belo da potarite i potrabilo novaz tamo plju ga madeze dotri.

All bills je i drug Julius

<u>CPI</u> is unstable

has an a Productive na many metric hadron energy investigation for the second second second second restriction of the second second second second second restriction of the second secon

uptierer fich. Zahrekan ur reng untieren migel (200 er fiche motte officieraing espeka, pomelas as and horn kontykka. U dessen justi par neuraliki, (signal birt ad jedenske keite. Gasenie je samakki, (signal je honom lažnog samopoutanja keji je take neu orens akvaložatema.

- Meni treba navar, gespeller Teta su tiro bih da mara kakui na vali salavi. Dve bili truta na tri masara Mulon da nure garantir. The first speed entering strategy dependent of the strategy of

a dekaju kota koja (n was adonad in span

SPANSKA STVARSONT I PRVI KORACI U NJUJ

> Grow in rule in he for local language A. Furbale

(pic source

1988-2002 - RISC & Moore's law

CPUs and Instruction set architecture (ISA) became simpler

But compiler grew to be more complex

RAM became larger and cheaper overall

RISC - Let's use fewer instructions, but it will be simple to figure out what happens inside

We've hit several 'walls' and started experimenting with ILP



What limits did we hit?

Power supply limit: more frequency -> more power -> more heat -> chip melts into magma

RAM limit - it's really slow compared to CPU registers

Speed of light: sending a signal through the whole big complex CPU is not an instant operation



Multiple Issue & Pipelining





Branch Prediction & Speculative Execution

Let's guess the result of a condition! (with about 70-80% probability!)

And <u>continue computing</u> instructions until proven guilty (or not!)



What did you say? Meltdown, Spectre? What's that?



ILP Saga

Register Renaming

Out-of-Order Execution

No-Lockup

Instruction Prefetching

Cliff Click - A Crash Course In Modern Hardware



2002 - ...

More <u>abstraction</u>

layers

Icons made by Freepik from www.flaticon.com

More co

Performance is dominated by

<u>cache misses</u>

(pic source)

Emery Berger - Performance Matters & V2!

Producing Wrong Data Without Doing Anything Obviously Wrong! (paper)

Mind the I

1 GB

3 GB

Automatic variables (local to a function's scope), caller's return address, etc. (grows towards lower memory addresses)

OS Kernel Space

User code cannot read from nor write to these addresses, otherwise resulting in a Segmentation Fault

ayout

Dynamic memory allocation through malloc/new free/delete (grows towards higher memory addresses)

> BSS Uninitialized static variables, filled with zeros

> > Data Static variables explicitly initialized

Text Binary image of the process (e.g., /bin/ls)

0×08048000

OXFFFFFFF

0xC0000000

pic sou

(pig

Hardware

- Both program and data it works with is located in RAM, just in different parts of it
- CPU runs instructions in an order, one by one sequentially
- Any program can ask CPU to read any memory address and do something with it

	0000000	push	ebp
	00000001	mov	ebp, esp
5	00000003	movzx	ecx, [ebp+arg_0]
	00000007	рор	ebp
	0000008	movzx	dx, cl
	00000000	lea	eax, [edx+edx]
	0000000F	add	eax, edx
1	00000011	shl	eax, 2
2	00000014	add	eax, edx
	00000016	shr	eax, 8
	00000019	sub	cl, al
	0000001B	shr	cl, 1
	0000001D	add	al, cl
-	0000001F	shr	al, 5
	00000022	MOVZX	eax, al
	00000025	retn	

(pic source

pic source

Hardware

hord up

00000000

00000001

0000000

000007

00000008

0000700

000.0

- Both program and data it works with is located in RAM, just in different parts of it
- CPU runs instructions in an order, one by 1000001 one sequentially
- Any program can ask CPT to read memory address and do venighing

Mait,

it

ebp bp.esp 30 ecx, [ebp+arg_0] .op dx, cl lea eax, [edx+edx] add eax, edx shl eax, 2 eax, edx eax, 8 cl, al cl, 1 al, cl al, 5 VZX eax, al tn

pic source)

pic source

von Neumann Architecture

Computer Systems - Von Neumann Architecture







ENTERPRISE AND THE CLOUD

BRENDAN GREGG



2ND EDITION

THINKING LOW LEVEL, WRITING HIGH LEVEL

RANDALL HYDE



O'REILLY'



THE REPORT OF TH



Micha Gorelick & Ian Ozsvald





Questions?

https://t.me/enchantner

O B B B B VOID

https://twitter.com/enchantner

https://www.linkedin.com/in/nickmarkov/

