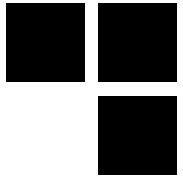
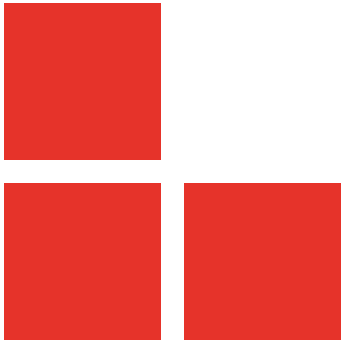
SYNACKTIV
DIGITAL SECURITY

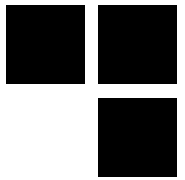


Program Optimization – introduction

simon.marechal@synacktiv.com

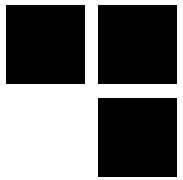
2023/2024





Introduction

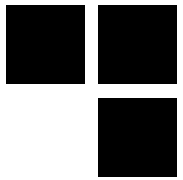
Course objective



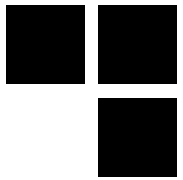
At the end of this course, you will have learned the following:

- Basic algorithmic complexity assessment;
- Heuristic solutions and their uses;
- A schematic view of how CPU and GPU work;
- How and what to profile;
- How to write GPU code.

What is program optimization?



It is the action of modifying a program to make it faster or more efficient, usually with regards to memory or power usage.

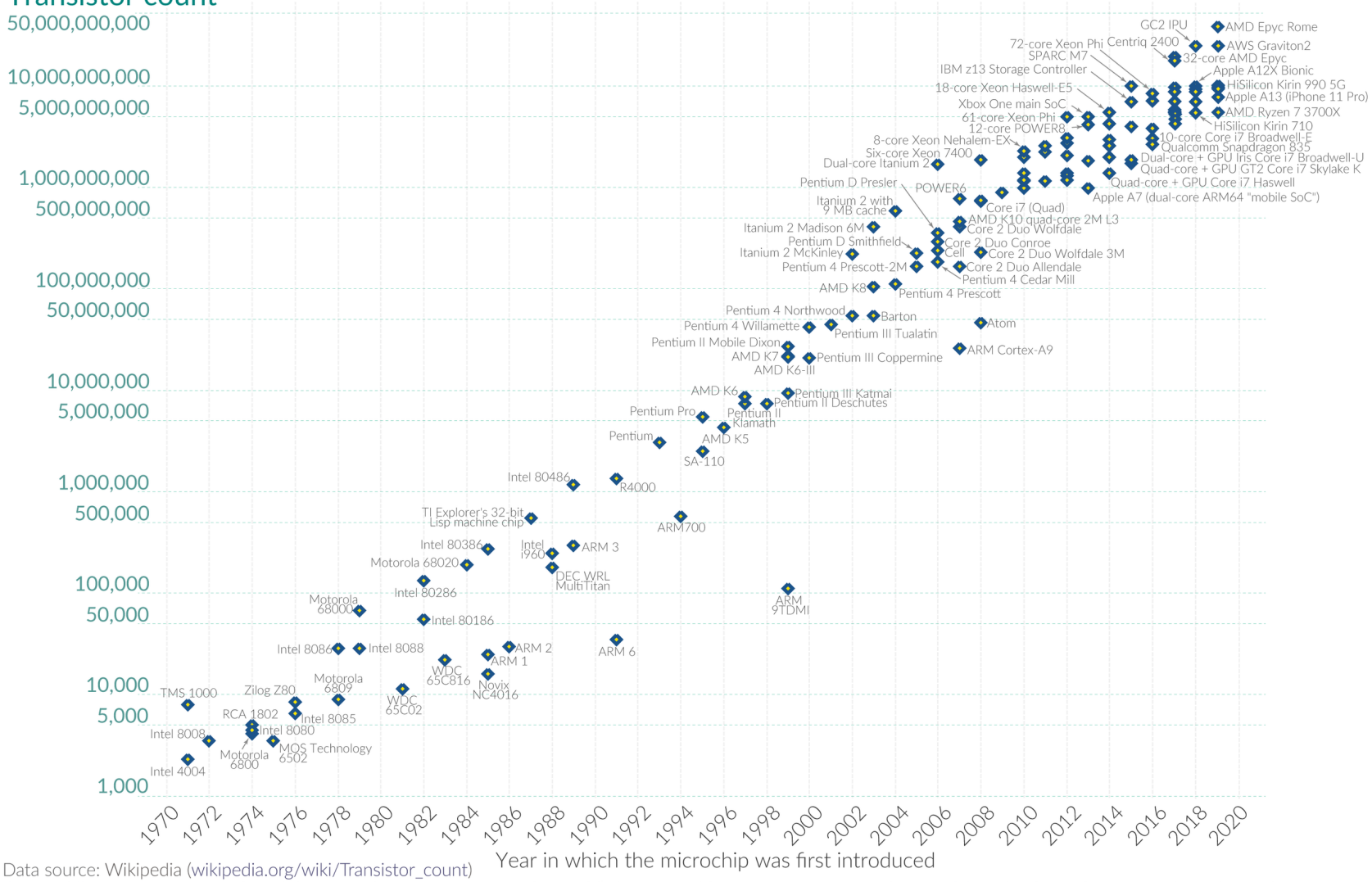


Why is it important ?

Moore's Law: The number of transistors on microchips doubles every two years

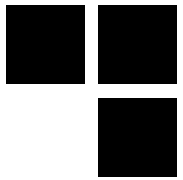
Our World in Data

Moore's law describes the empirical regularity that the number of transistors on integrated circuits doubles approximately every two years.
This advancement is important for other aspects of technological progress in computing – such as processing speed or the price of computers.



Data source: Wikipedia (wikipedia.org/wiki/Transistor_count)
OurWorldinData.org – Research and data to make progress against the world's largest problems.

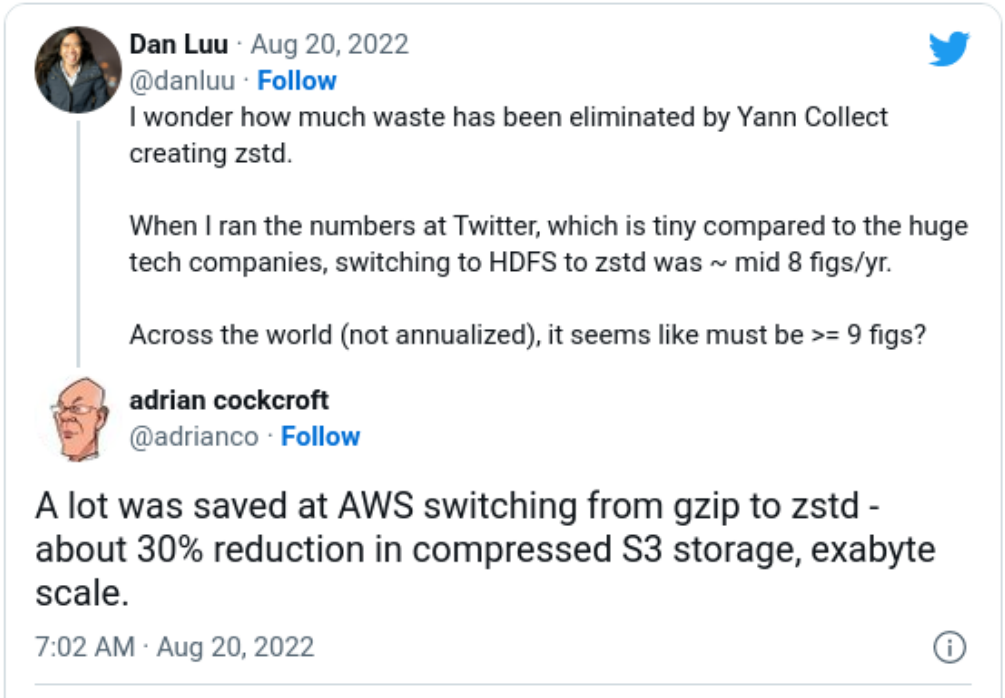
Nothing is optimized, nothing is fast



- thanks for Moore's law, people expect free speed ups over time ;
  - but computers do not feel faster!
- dynamic languages have become hugely popular ;
- people don't really care about performance when performance is good enough.

However ...

- small speedups make a huge difference at scale



Dan Luu · Aug 20, 2022
@danluu · **Follow**

I wonder how much waste has been eliminated by Yann Collect creating zstd.

When I ran the numbers at Twitter, which is tiny compared to the huge tech companies, switching to HDFS to zstd was ~ mid 8 figs/yr.

Across the world (not annualized), it seems like must be >= 9 figs?

adrian cockcroft
@adrianco · **Follow**

A lot was saved at AWS switching from gzip to zstd - about 30% reduction in compressed S3 storage, exabyte scale.

7:02 AM · Aug 20, 2022

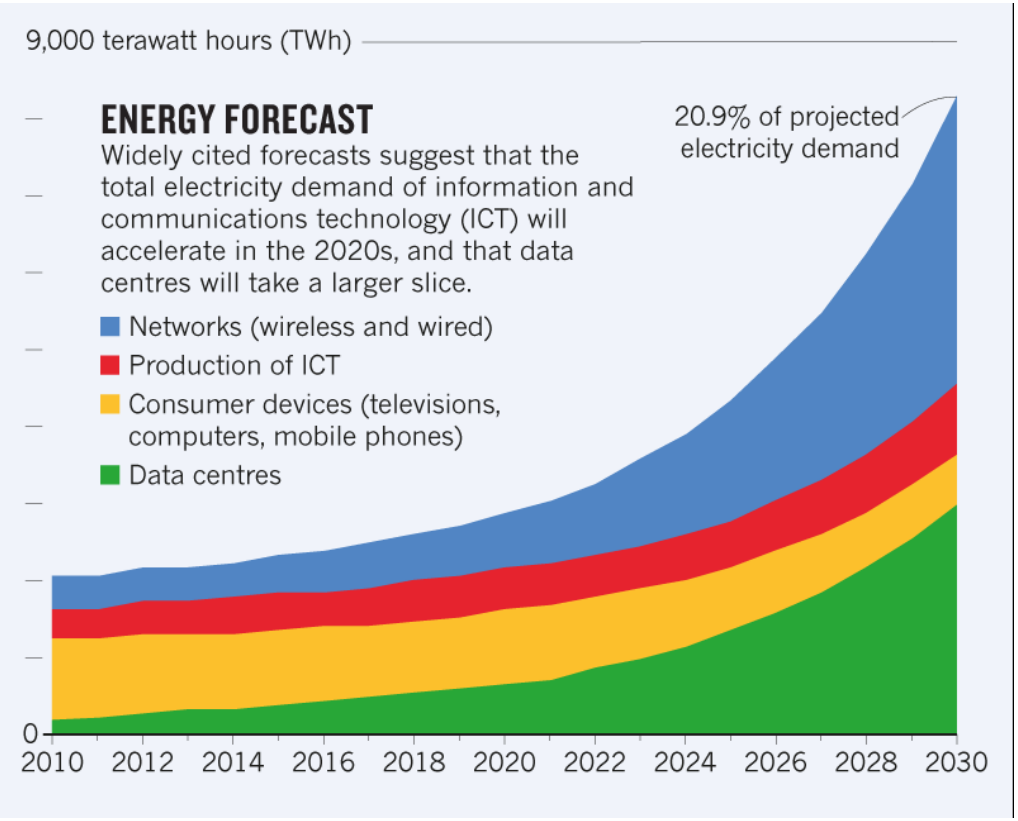
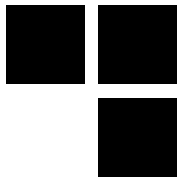
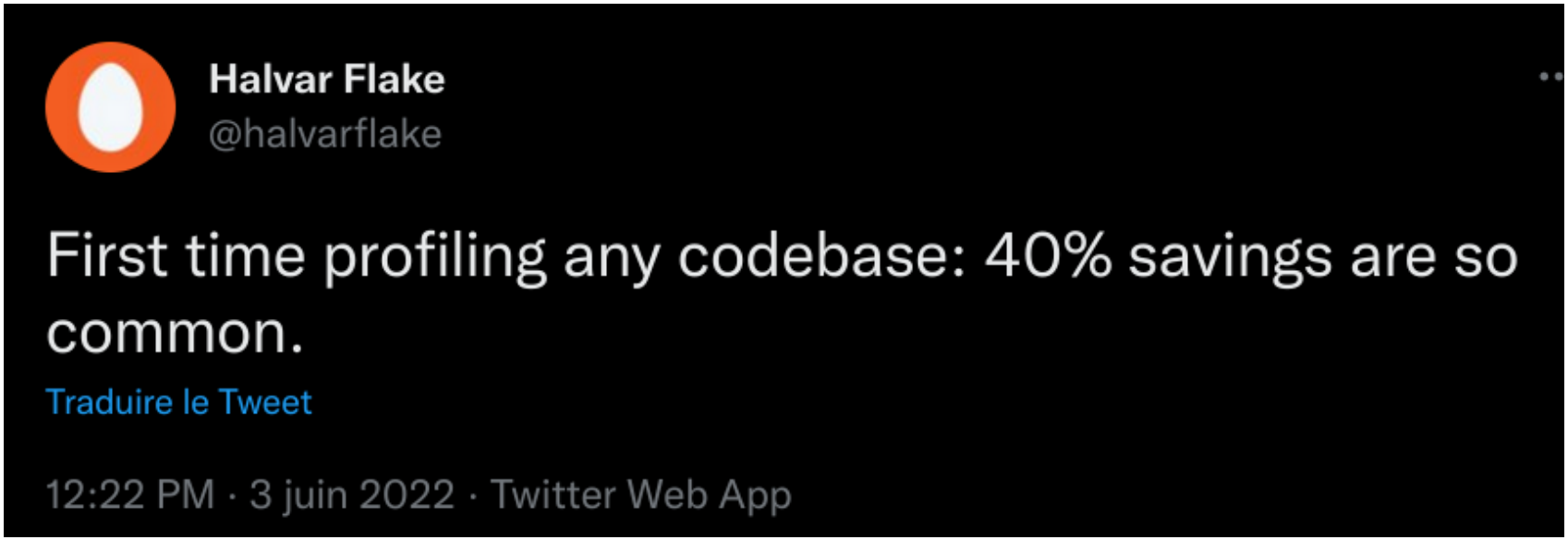


Figure 1: Source: The Real Amount of Energy A Data Center Uses - Clarissa Garcia

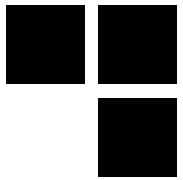
It is usually easy!



- Nothing is optimized !



A rough idea of languages performance

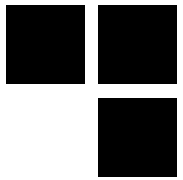


In terms of computing speed:

- low level languages (C, C++, Rust),
- static languages (.NET, Java, Go, Haskell, Swift), around 3x/5x slower
- JavaScript, 5x/10x slower
- dynamic languages (Python, Ruby), 10x/50x slower

In theory, JIT languages could be the fastest, but this is not the case in practice

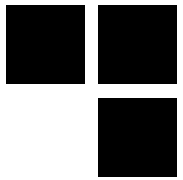
Is my program fast?



CPU tick	0.3ns	1s
L1 cache	1.18ns	4s
L2 cache	4.1-6.5ns	14-22s
L3 cache	15.9-16.5ns	54-56s
64 bits division	10-26ns	35-88s
RAM access	89-91ns	~5 minutes
NVMe access	300us	~11.5 days
Same LAN, ping	0.5ms	~19.5 days
Container boot	0.5-2s	52-211 years
Dell server reboot 😠	2 minutes	12675 years

Figure 2: i7 6700 - DDR4-2666

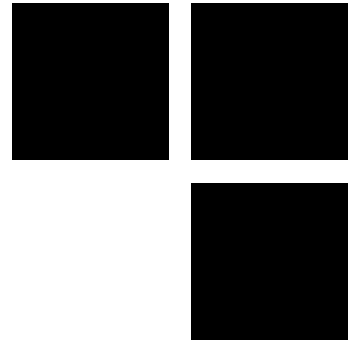
Is my GPU program fast?



GPU tick (GTX 1080)	0.6ns	1s
Cache ?	22.4ns	37s
Memory latency	244ns	1m46

Figure 3: GTX 1080

QUESTIONS?



Thank you for your attention

SYNACKTIV
DIGITAL SECURITY