

What you can do with Dask

- Big Pandas
- Parallel For Loops
- Big Arrays
- Machine Learning

Big Pandas

Dask DataFrames use pandas under the hood, so your current code likely just works. It's faster than Spark and easier too.

[Documentation](#) [Performance Benchmarks](#)

$n = 10 \rightarrow$ distributed computing

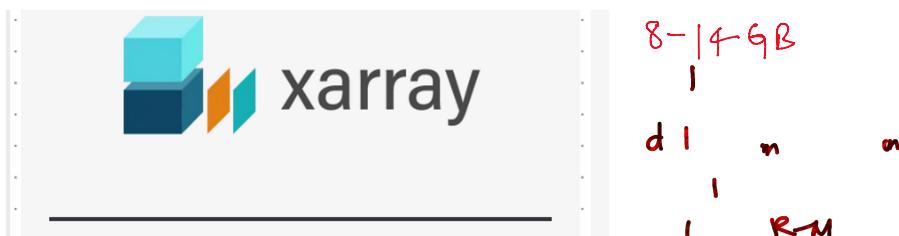
```
import dask.dataframe as dd

df = dd.read_parquet("s3://data/uber/")

# How much did NYC pay Uber?
df.base_passenger_fare.sum().compute()

# And how much did drivers make?
df.driver_pay.sum().compute()
```

<https://www.dask.org/>



National Water Model

plane operations
↓
(16GB RAM) ??

Tools

Xarray Zarr Flox

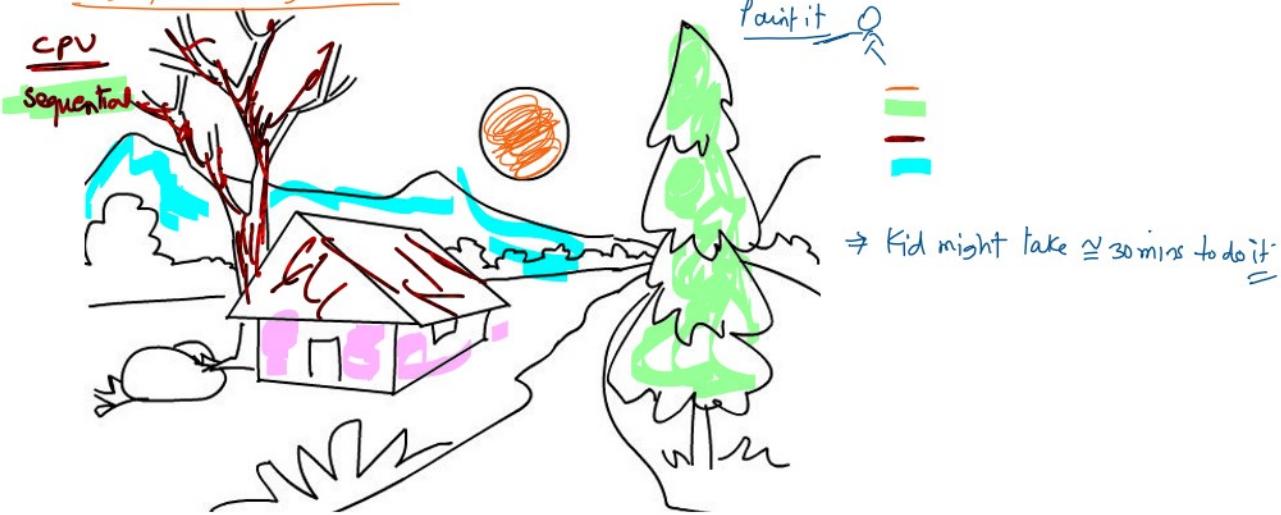
Data

250 TB of Zarr data → Ignore
not for us

Introduction to Graphics Processing Units (GPUs)

Acceleration

Page from drawing book

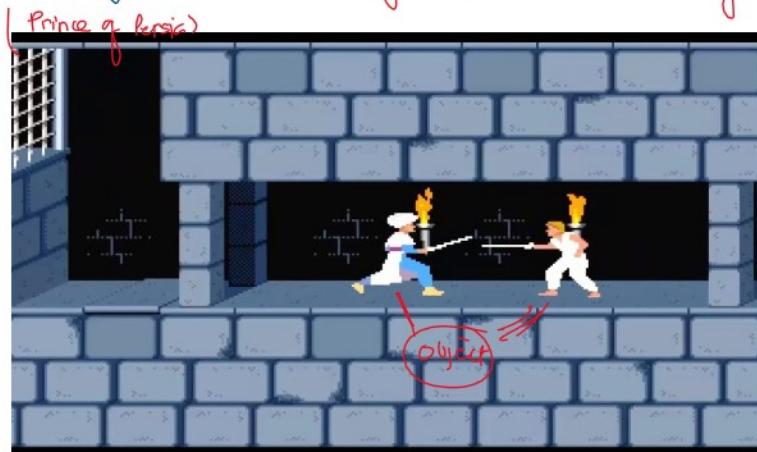


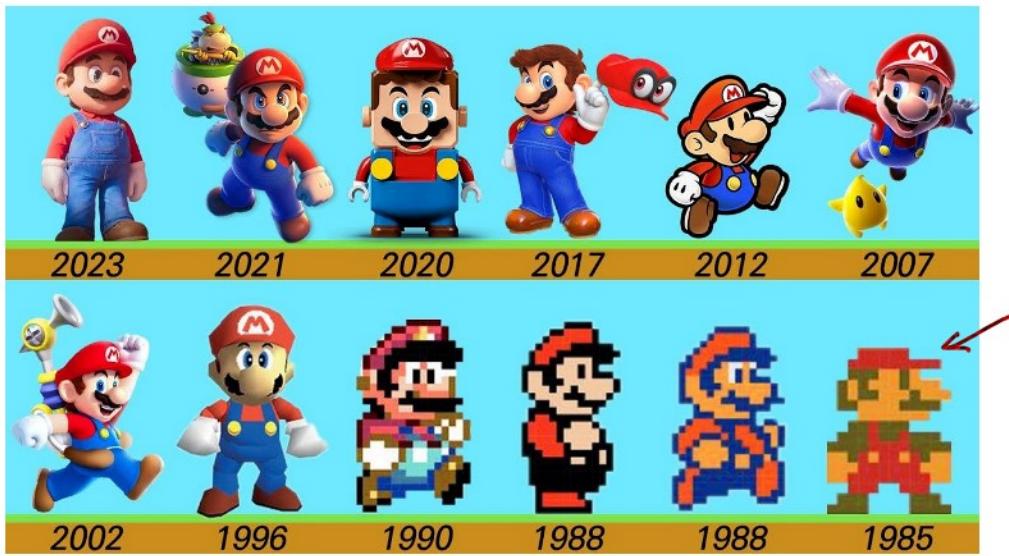
GPU 4-5 kids @ home doing painting for pizza/KFC burger





GPGUs are highly parallel processors originally designed for rendering graphics. They are now widely used for accelerating deep learning computations due to their ability to perform many operations simultaneously.





CPU : central processing unit

- general purpose processor

Purpose documentation, audio | video playing, browsing, etc

- Modern CPUs: 2 to 8 cores

Make 60 mins video

with a lot of editing

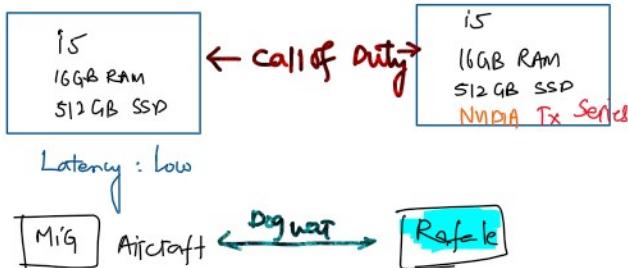
- Modern CPUs: 2 to 8 cores

Make 60mins video
with a lot of editing

Gamer: Srinivas HS.

CAMTASIA

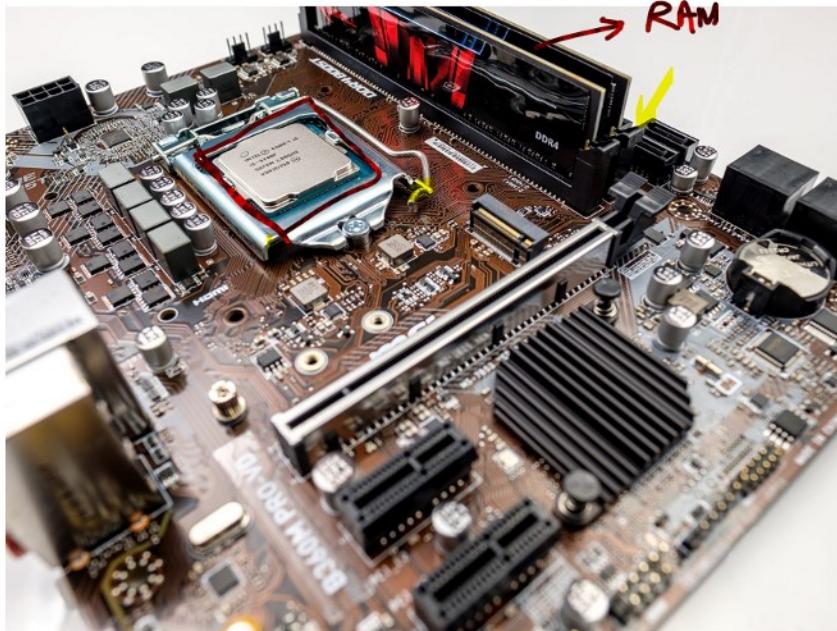
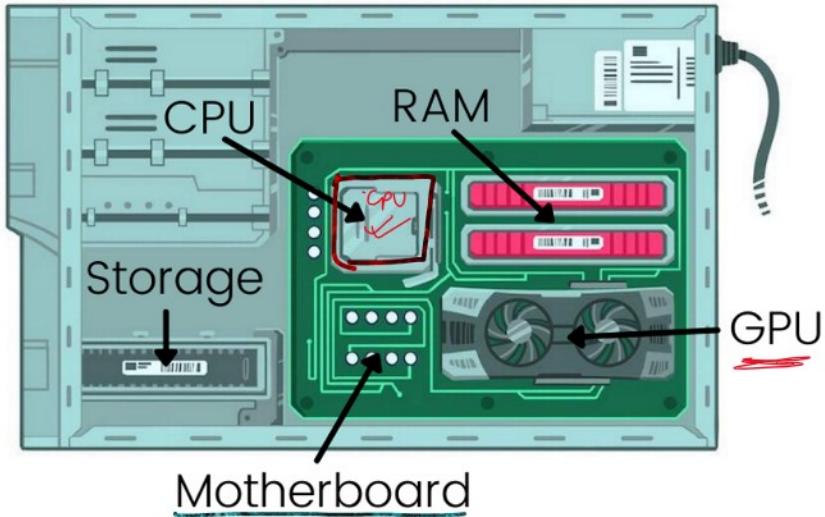
Gamer: Francis

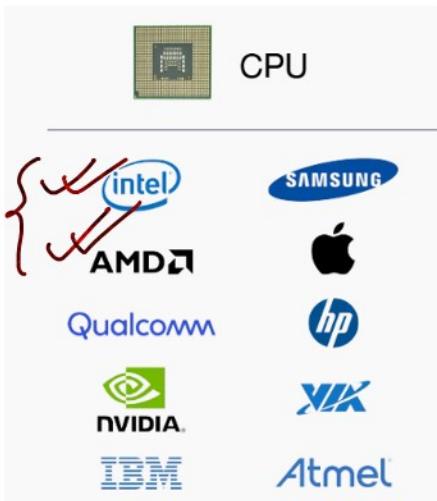


CPU Features Summary:

- Has Several Cores
- Low Latency
- Specialized in Serial Processing
- Capable of executing a handful of operations at once
- Have the highest FLOPS utilization for RNNs (recurrent neural network)
- Support the largest model thanks to its large memory capacity
- Much more flexible and programmable for irregular computations (e.g., small batches non MatMul computations)



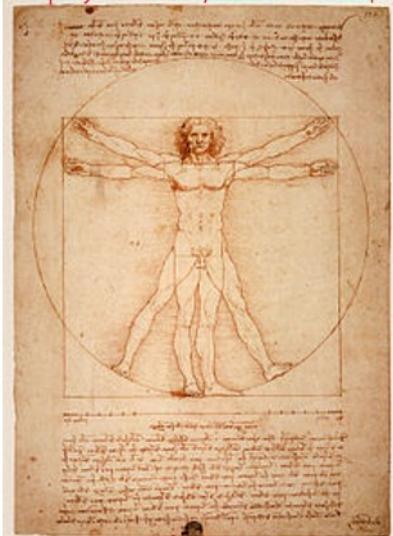




GPU : Graphics Processing Unit

- is a specialized processor which works as a performance accelerator with the CPU
- In comparison to CPU, GPU has thousands of cores → it helps to break down complex problems into thousands or millions of separate tasks to perform computations in parallel.

Medical
Space
CT / MRI
+ domain
→ Philips
→ GE
+ many more.



Leonardo da Vinci ✓



Leonardo da Vinci made the first real studies of flight in the 1480s. He had over 200 drawings and sketches that illustrated his theories on flight. His ornithopter flying machine was an aircraft that would fly by flapping its wings, a design he created to show how humans could fly. It even had a sophisticated flight control system; however, this design was never built by the designer.

<https://sandiegoairandspace.org/collection/item/leonardo-da-vinci-ornithopter-mock-up#:~:text=Leonardo%20da%20Vinci%20made%20the,show%20how%20humans%20could%20fly.>

<https://airandspace.si.edu/stories/editorial/leonardo-da-vinci-and-flight>

Graphics Processing

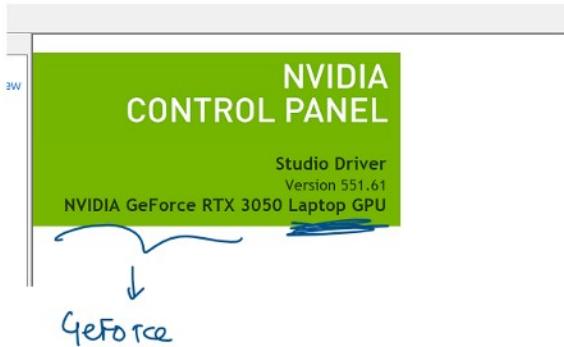
Video Rendering

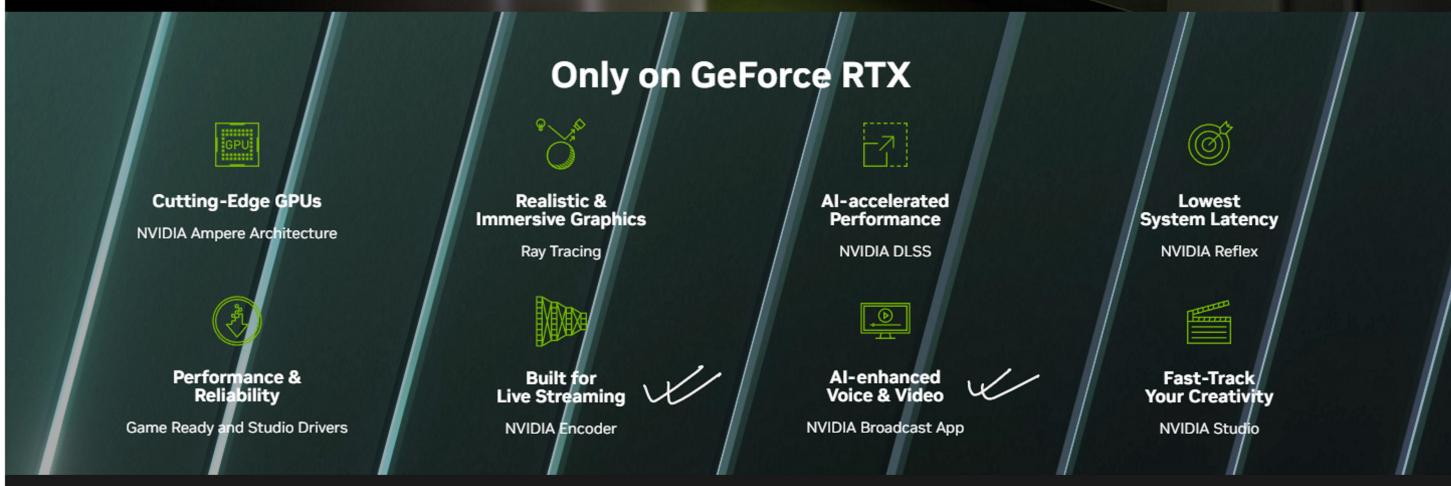
ML

Mining cryptocurrencies like Bitcoin ↴

GPU Features Summary:

- Has thousands of cores
- High throughput
- Specialized for parallel processing
- Capable of executing thousands of operations at once





'Largest Graphics card'



[NVIDIA GTC 2020 Keynote in 10 minutes: Updated RTX and A100 GPU system](#)



T: Tensor Processing Units

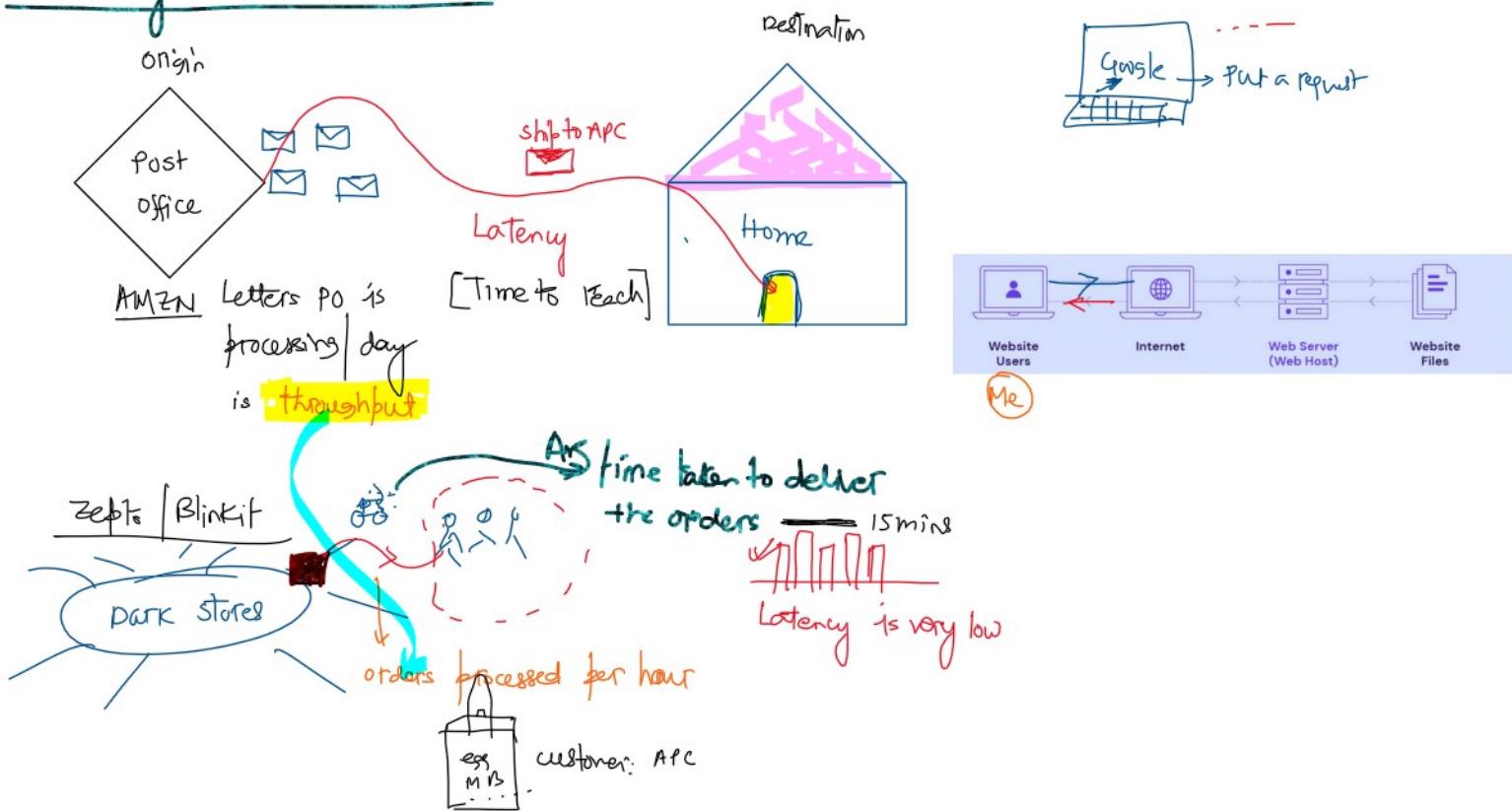
- designed and built in 2015
- Made that public 2018
- custom developed application-specific integrated circuits (ASICs)

- Made that public 2018
- custom developed application-specific integrated circuits (ASICs) primarily for accelerating ML and DL workloads

Key benefit:

- optimized for **TensorFlow** - designed to work seamlessly with TF
- **Energy efficiency** - Thus are more energy-efficient for certain types of neural network operations.
- **Training Large Models** - ideal for training very large models basically can run hourly forecasts.

Latency vs Throughput



High throughput and Low Latency.

Note: In general, latency is closely associated with CPU speed and throughput is closely associated with memory (RAM).

• In general, latency is closely associated with CPU speed and throughput is closely associated with memory (RAM)

① Device specifications

Device name	BRUCE
Processor	11th Gen Intel(R) Core(TM) i7-11800H @ 2.30GHz 2.30 GHz
Installed RAM	16.0 GB (15.8 GB usable)
Device ID	F3E5F38B-5CEB-438A-AC41-814A38BF0397
Product ID	00327-35939-70884-AAOEM
System type	64-bit operating system, x64-based processor
Pen and touch	Pen support



GPUs benchmarks hierarchy 2024

https://www.tomshardware.com/reviews/gpu-hierarchy_4388.html

Intel's market share is growing again, says Mercury Research

Mobile CPU share	2024 Q1 current quarter share	Share change (points) year
Intel	80.7%	- 3.1
AMD	19.3%	+3.1
Total	100.0%	

10 May 2024

OS