

Estimate Cost of Training LLMs from Scratch

Instructor

Sourab Mangrulkar

Machine Learning Engineer at

Creator of PEFT



Primary Factors

- No. of parameters
- No. of training tokens
- Hardware

How to estimate
the cost?

1

Computational Cost

How to estimate the cost?

- 1 Computational Cost
- 2 Hardware cost

Computational Cost

No. of floating point operations (FLOPS)

Computational Cost

- Depends on the no. of model parameters and training tokens

Computational Cost

- Depends on the no. of model parameters and training tokens
- $6N$ FLOPS/Token

Computational Cost

- Depends on the no. of model parameters and training tokens
- $6N$ FLOPS/Token
- Llama 2

Computational Cost

- Depends on the no. of model parameters and training tokens
- $6N$ FLOPS/Token
- Llama 2
No. of parameters: 70B

Computational Cost

- Depends on the no. of model parameters and training tokens
- $6N$ FLOPS/Token
- Llama 2
 - No. of parameters: 70B
 - No. of tokens: 2T

Computational Cost

- Depends on the no. of model parameters and training tokens
- $6N$ FLOPS/Token
- Llama 2
 - No. of parameters: 70B
 - No. of tokens: 2T
 - Computational Cost: **$8.4e23$ FLOPS** ($6 \times 7e10 \times 2e12$)

How to estimate the cost?

- 1 Computational Cost
- 2 Hardware cost

Hardware Cost

- Depends upon the type of GPU/TPU

Hardware Cost

- Depends upon the type of GPU/TPU
A100: ~3.12e14 FLOP/s

Hardware Cost

- Depends upon the type of GPU/TPU
A100: ~3.12e14 FLOP/s
- Model Flops Utilisation(MFU)

Hardware Cost

- Depends upon the type of GPU/TPU
A100: ~ $3.12e14$ FLOP/s
- Model Flops Utilisation(MFU)
MFU: ~ $1.5e14$ FLOP/s

Hardware Cost

- Depends upon the type of GPU/TPU
A100: ~3.12e14 FLOP/s
- Model Flops Utilisation(MFU)
MFU: ~1.5e14 FLOP/s
- Time Taken

Hardware Cost

- Depends upon the type of GPU/TPU
A100: $\sim 3.12e14$ FLOP/s
- Model Flops Utilisation(MFU)
MFU: $\sim 1.5e14$ FLOP/s
- Time Taken
No. of hours: $8.4e23 / (1.5e14 * 3600) = 1.55e6$ A100 hours

Hardware Cost

- Depends upon the type of GPU/TPU
A100: $\sim 3.12e14$ FLOP/s
- Model Flops Utilisation(MFU)
MFU: $\sim 1.5e14$ FLOP/s
- Time taken
No. of hours: $8.4e23 / (1.5e14 * 3600) = 1.55e6$ A100 hours
- Total Cost
Cost: $1.55e6 * (\$1.8/\text{hr}) \sim \$3M$

Cost of Training LLMs from Scratch

Cost of training LLMs

- $(\text{Computational Cost/MFU}) * \text{Hardware Cost}$

Thank You
