

# NVIDIA H100 VS L40S: UNLEASH THE BEAST OR MASTER THE BALANCE?

The AI revolution is here, and selecting the perfect tool is paramount. NVIDIA's H100 and L40S stand at the pinnacle of GPU technology, each excelling in its own domain. Let's explore their strengths to assist you in picking out the ideal solution for your requirements.

## H100: THE UNYIELDING FORCE. PICTURE A TITAN, FORGED IN THE CRUCIBLE OF UNPARALLELED COMPUTATIONAL MIGHT. THE H100 BOASTS:

- **Unrivalled Tensor Core performance:** Dominate AI training and inference with 8 times the throughput of the A100
- **Lightning-fast memory bandwidth:** Indulge in a feast of data at 5 TB/s, empowering the most demanding models
- **State-of-the-art interconnects:** Establish rapid communication between GPUs, obliterating performance barriers

## THE H100 IS THE ULTIMATE CHOICE FOR:

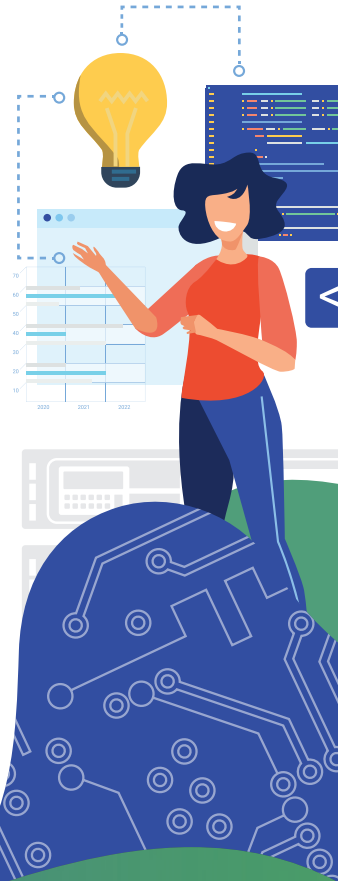
- Pushing the boundaries of AI research and development
- Handling massive, intricate models with demanding memory needs
- Constructing server clusters that redefine performance standards

## L40S: THE VERSATILE MAESTRO. PICTURE IT AS A SKILLED FENCER, EFFORTLESSLY BLENDING POWER WITH FINESSE. THE L40S SHINES WITH:

- **Balanced performance:** Excel in both AI and graphics tasks, from deep learning to rendering stunning visuals
- **Cost-effective efficiency:** Achieve remarkable performance without the H100's substantial price tag
- **Broader availability:** Say goodbye to waiting – the L40S is readily accessible to fuel your projects

## THE L40S IS IDEAL FOR:

- Implementing AI solutions in production, from chatbots to medical imaging
- Creative professionals requiring both AI acceleration and graphics capabilities
- Businesses in search of a cost-effective entryway into high-performance computing



Ultimately, the decision is yours. Do you crave the sheer power of the H100 to conquer the AI frontier? Or do you lean towards the balanced elegance of the L40S, a master of both realms? Whichever path you select, remember, with NVIDIA, you're always ahead of the curve.

## DATA CENTRE GPU WORKLOAD COMPARISON

GPU MODEL	DL TRAINING & DA	DL INFERENCE	HPC / AI	OMNIVERSE / RENDER FARMS	VISUAL WORKSTATIONS	VIRTUAL DESKTOP / VDI	MAINSTREAM ACCELERATION	FAR EDGE ACCELERATION	AI ON 5G
A100	✓	✓	✓				✓		✓
H100	✓✓	✓✓	✓✓				✓✓		
L40 S		✓✓		✓✓	✓✓	✓	✓✓	✓	

✓ GOOD    ✓✓ BEST

## DATA CENTRE GPU COMPARISON

GPU MODEL	DESIGN	FORM FACTOR	MAX POWER	FP 64 TC   FP32	TF32 TC   FP16	FP8 TC   INT8	MEMORY	VGU
A100 High Perf	Compute	SXM4	500W	19.5   19.5	312   624	NA   1248	80GB HBM2e	YES
		x16 PCIe Gen4 2 Slot FHFL 3 NVLink Bridge	300W	19.5   19.5	312   624	NA   1248	80GB HBM2e	
H100	Highest Perf AI, Big NLP, HPC, DA	SXM5	700 W	67   67	989   1976	3958   3958	80 GB	YES
		X16 PCIe 2 Slot FHLL 3 NVLnk Bridge	350 W	51   51	756   1513	3026   3026	80 GB	YES
		X16 PCIe Dual 2 Slot FHLL 3 NVLnk Bridge	2 X 400 W	134   134	1976   3958	7916   7916	188 GB	YES
L40 S	Powerful Graphics + AI	x16 PCIe Gen4 2 Slot FHFL	300W	NA   TBD3	TBD <sup>3</sup>   TBD <sup>3</sup>	TBD <sup>3</sup>   TBD <sup>3</sup>	48GB GDDR6	YES

## IF YOU'D LIKE TO FIND OUT MORE, PLEASE GET IN TOUCH:

### SOUTH AFRICA

**E:** sales@boston-it.co.za

**T:** (+27) 11 014 1924

**W:** www.boston-it.co.za

Follow us: Keep up to date with Boston on socials!

**LinkedIn:** linkedin.com/company/boston-it-solutions-africa

**X:** twitter.com/bostonlimited

