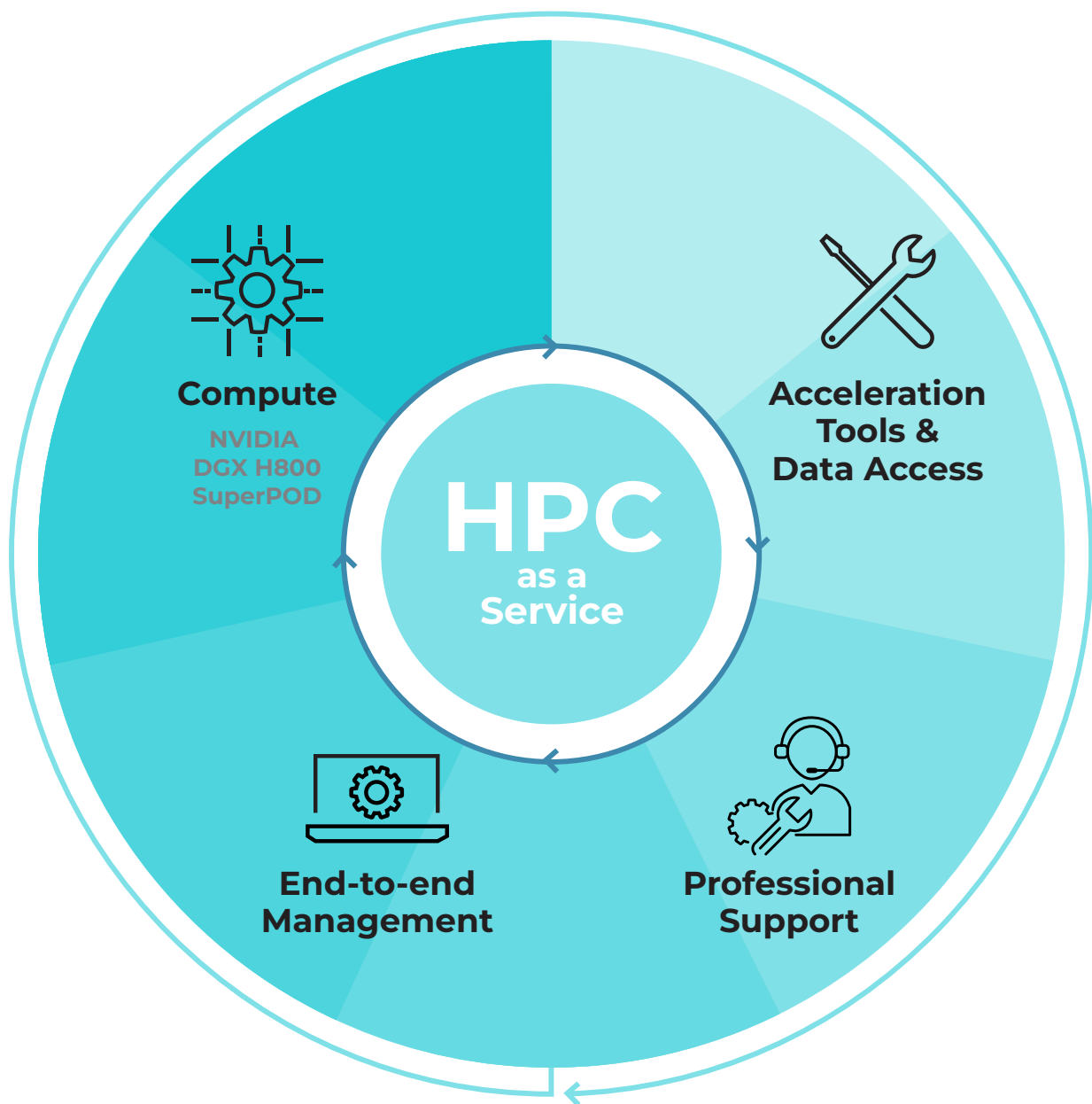# HKSTP

High Performance Computing

as a Service

SERVICE OVERVIEW

With HKSTP High-Performance Computing as a Service (HPC as a Service), you can innovate faster and make every dollar count.

Our comprehensive services are powered by top-notch computing, acceleration tools, and data access, providing a robust foundation to accelerate your development cycle.

Beat Time to Market

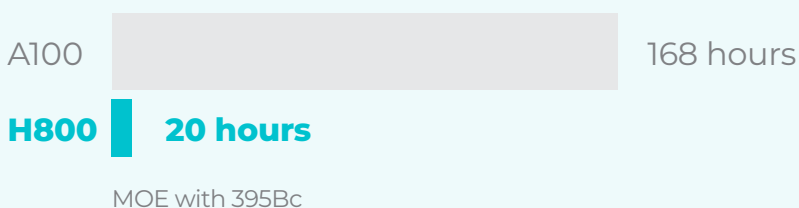Our services deliver top-tier computing for training, inference, and the most challenging AI workloads.

This compute foundation offers unparalleled performance and adaptability to accelerate data science pipelines.
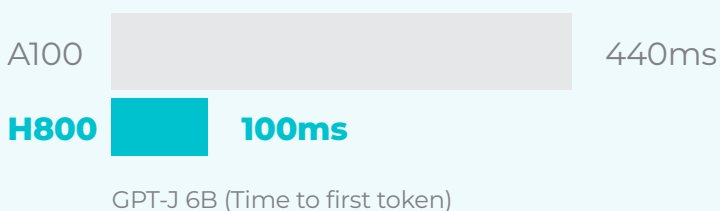
9X
FASTER

How fast is H800? Let's compare:

9X
faster #

Fastest Time to Train on Every Workload

A100 — 168 hours

H800 — **20 hours**

MOE with 395Bc

4.4X
faster #

Enhanced Real-Time Inference Capability

A100 — 440ms

H800 — **100ms**

GPT-J 6B (Time to first token)

Source: Nvidia
The results are dependent on simulation and displayed for reference only.

Best Value of Investment

GPU's parallel processing capability splits complex computing tasks into thousands smaller tasks that run simultaneously. This keeps your resources and team busy, not idle or wasted.

Maximise your investment value when you can expect faster results in limited time.

75%
COST SAVED

How does it make sense from a financial standpoint?

We can use MosaicML's analysis on training a 7B parameter LLM with 134B tokens.

GPU	GPU Hours to Train	Approx. Cost (HKD)	
8 x A100	11,462	1,283,744	
8 x H800	**2,000**	**320,000**	**75% cheaper**

FlashAttention2 Training on a 175B LLM

The figures for A100 references Ori, while the estimated cost for H800 is HKD20 per hour.
The results are dependent on simulation and displayed for reference only.

EXPERIENCE

End-to-End Management

Time is luxury in R&D. Your focus should be on core tasks, not cumbersome operations.

SHIP and GO

What We Offer

1 Security

2 Multitenancy Architecture

3 Acceleration Tools

4 Full-stack Environment

What You Get

- Secure and fast data transfer through a local data center managed and operated by HKSTP.

- GPU re-allocation or scale on demand at ease.

- Pre-trained AI models, software tools, and cross-industry data access.

- Workload performance intelligence, software licensing, compilers and libraries

PLANS & PRICE

We offer a straightforward bill rate.

This pricing transparency ensures you know the exact costs and adjust resources as you grow.

H800 Resource	Daily (HKD)	Monthly (HKD)	Annual (HKD)
1/4 card	-	$3,600	-
1/2 card	-	$7,200	-
1 card	$1,440	$14,400	-
2 cards	$2,880	$28,800	-
3 cards	$4,320	$43,200	-
4 cards	$5,760	$57,600	-
5 cards	$7,200	$72,000	-
6 cards	$8,640	$86,400	-
7 cards	$10,080	$100,800	-
8 cards	$11,520	$115,200	-
Dedicated DGX (8 cards)	-	-	**Contact Us**

FREE PLAN **1 Month** **FREE PLAN**

Try 1 Month FREE

✓ Onboarding support
✓ Consultation services
✓ Park companies ONLY; nomination is required

* Subject to Terms and Conditions.
* 1/8 H800 card is allocated for Free Plan users.
* Please contact your account manager for application.

HKSTP

📍 Service Centre,
2/F, Building 19W
Hong Kong Science Park

📞 2629 6736

✉️ DSH@hkstp.org

BOOK CONSULTATION