

TECHNICKÁ SPECIFIKACE DODÁVKY

Grafický výpočetní server pro modelování kapalných fází

Popis zařízení: Předmětem dodávky je autonomní výpočetní systém s GPU čipy pro simulace vícefázového proudění na bázi SPH metody, HPC simulací v oblasti výpočtů z prvních principů a pro simulace v oblasti machine learning a umělé inteligence.

Požadavky na hardware:

- Výpočetní systém musí sestávat z právě jednoho výpočetního serveru osazeného výpočetními GPU kartami, které jsou propojeny vysokorychlostní komunikační sítí.
- Výpočetní systém musí poskytovat agregovaný výpočetní výkon minimálně 2,5 Pflop/s v režimu FP16 (half precision floating point).
- Výpočetní systém musí poskytovat vysoce propustnou operační paměť (HBM) GPU karet o celkové velikosti minimálně 320 GB.
- Výpočetní server výpočetního systému musí splňovat následující požadavky:
 - Architektura x86-64
 - Jeden procesor na výpočetní server
 - Minimálně 64 fyzických CPU jader a 128 vláken na procesor serveru
 - Teoretický výpočetní výkon procesoru serveru minimálně 2,3Tflop/s v režimu dvojité přesnosti (double precision floating point)
 - Kapacita operační paměti RAM serveru minimálně 512 GB DDR4, min. 3200MT/s
 - Sběrnice PCI Express generace 4.0
 - Jeden lokální disk technologie NVMe SSD o kapacitě minimálně 1,92TB (pro operační systém)
 - Lokální disk technologie NVMe SSD o celkové kapacitě minimálně 7,68TB a podporou rychlosti náhodného čtení 1,4 milionu IOPS (pro data)
 - Minimálně 2 porty 10Gbase-T Ethernet
 - Minimálně 1 port 1Gbase-T Ethernet dedikovaný pro management systému
 - Trusted Platform Module (TPM 2.0)
- Výpočetní systém musí obsahovat minimálně 4 GPU karty shodného typu určené pro výpočty a alespoň 1 GPU kartu určenou pro vizualizaci výsledků výpočtů.
- Každá GPU karta určená pro výpočty musí splňovat následující požadavky:
 - Podpora technologie CUDA
 - Teoretický výpočetní výkon minimálně 312TFLOPS v režimu FP16 (half precision floating point)
 - GPU paměť minimálně 80 GB HBM2
 - Propustnost GPU paměti minimálně 2,0 TB/s
- Karty pro výpočty musí být vzájemně propojeny vysokorychlostní GPU sítí.
 - GPU síť musí umožňovat současnou komunikaci libovolných GPU karet mezi sebou.

- Obousměrná šířka pásma mezi libovolnými dvěma GPU moduly musí být alespoň 600GB/s.
- Topologie GPU sítě musí být plně neblokující.
- Karta určená pro vizualizaci musí splňovat následující požadavky:
 - Minimálně 4 Mini DisplayPorty
 - podpora rozlišení 5k,
 - minimálně 4GB GDDR6 GPU paměti
- Všechny použité komponenty musí být vzájemně plně kompatibilní a musí být zaručena jejich 100% funkcionality.
- Výpočetní systém může být v budoucnu sdílen mezi více vědeckými týmy a přesouván mezi lokalitami, proto požadujeme provedení skříně typu tower a bezúdržbové interní chlazení výpočetního systému dimenzované pro provoz v prostředí bez aktivního chlazení. Systém bude provozován v prostředí vědeckých pracovišť, proto jeho hlučnost při plném výkonu nesmí přesáhnout hladinu 37dB a zároveň jeho maximální kapacita napájecích zdrojů nesmí přesáhnout 1550W / 220V.
- Dodavatel musí v nabídce uvést detailní konfiguraci hardware výpočetního systému – výpočetního serveru, GPU karet a GPU sítě.

Požadavky na software:

- Součástí dodávky musí být veškeré potřebné programové vybavení a licence potřebné pro legální použití výpočetního systému v plném rozsahu zadavatelem a uživateli služeb zadavatele. Programovým vybavením systémů se rozumí veškeré softwarové vybavení – firmware, ovladače hardware, operační systém, knihovny, softwarové stacky, aplikační software, management software, atd. výpočetního systému.
- Výpočetní server výpočetního systému musí používat operační systém typu Linux, 64-bitový, nativně používající instalační balíky typu RPM nebo DEB.
- Použitý operační systém musí mít dobrou dostupnost aktualizací, zejména těch, které jsou určeny ke zvýšení bezpečnosti a spolehlivosti.
- Výpočetní systém musí podporovat běh následujících open-source frameworků pro strojové učení a umělou inteligenci na GPU kartách systému:
 - TensorFlow
 - PyTorch
 - Keras
- Výpočetní systém musí umožňovat běh kontejnerů (tj. použití virtualizace na úrovni operačního systému) typu Docker. Kontejnery musí podporovat GPU karty.
- Výpočetní systém musí poskytovat transparentní přístup do GPU pamětí GPU karet mezi GPU kartami prostřednictvím GPU sítě.
- Dodavatel musí v nabídce uvést název operačního systému a popis software výpočetního systému.

Požadavky na instalaci:

- Dodavatel musí výpočetní systém (hardware a software) nainstalovat a zprovoznit v infrastruktuře zadavatele. Výpočetní systém bude instalován a provozován v prostorách zadavatele v budově VTP C1. Budova VTP C1 se nachází na adrese Teslova 5b, 301 00 Plzeň.
- Jelikož se jedná o vysoce specializovaný GPU výpočetní systém, součástí dodávky musí být odborné zaškolení technickým specialistou s oficiální certifikací pro instalaci nabízeného systému v délce jednoho dne v prostorách zadavatele.
- Instalací do datového centra zadavatele se rozumí veškeré dodávky a činnosti, jejichž výsledkem bude zprovoznění výpočetního systému v prostorách NTC.
- Instalace do prostor zadavatele musí být provedena dle platné legislativy a předpisů a dle požadavků a doporučení výrobců zařízení.

Požadavky na záruku:

- Výpočetní systém musí být pokryt originální zárukou výrobce v minimální délce 3 roky, s výměnou vadného dílu on-site do 3 dní a kontaktní linkou dostupnou v režimu 24/7.
- Záruka výrobce musí pokrývat všechny komponenty tohoto systému a musí zahrnovat i softwarovou podporu výpočetního systému včetně operačního systému, GPU ovladačů a podporovaných open-source AI frameworků.