**Příloha č. 1 zadávací dokumentace**

**Technická dokumentace - specifikace požadovaného plnění**

**CESNET -** **Dodávka HD výpočetních clusterů pro MetaCentrum (2022)**

1. **1.**  **Základní informace a požadavky na výpočetní uzly**
	1. Předpokládaná a zároveň pevná (fixní) cena této veřejné zakázky je 20 700 000,- Kč bez DPH.
	2. **Místem plnění je**:
	* Masarykova univerzita, Ústav výpočetní techniky, Botanická 68a, 602 00 Brno (dále jen „**MU**“)
	* Ústav molekulární genetiky AV ČR, Vídeňská 1083, 142 20 Praha 4 (dále jen „**ÚMG**“), Budova F, suterén, místnost 01.129.1 - Datové centrum Z3
	1. **Obecné požadavky**
		1. **Předmětem** plnění veřejné zakázky je kompletní řešení, sestávající se z dodávky, instalace a zprovoznění alespoň **50 identických uzlů (serverů) výpočetních clusterů** podle specifikace uvedené dále. *(Identickým uzlem* se rozumí osazení zcela shodných komponent do každého uzlu, včetně použitých pamětí; výjimkou může být počet volných PCIe slotů, viz odst. 1.4.12).
		2. Sestavy musí být umístěné v lokalitách tak, aby prvních 40kW maximálního příkonu strojů (zaokrouhleno nahoru na celá šasí) bylo umístěno v lokalitě ÚMG, zbytek uzlů v lokalitě MU.
		3. Součástí plnění je rovněž dodávka a instalace potřebné síťové infrastruktury zahrnující kabeláž. Switche nejsou součástí dodávky (vyjma případných switchů pro IPMI management, viz odst. 1.4.18).
		4. Součástí předmětu plnění je rovněž poskytnutí záruky za jakost a řádnou funkčnost dodaného plnění, včetně technické podpory (dále jen „záruka“), na dobu nejméně 36 měsíců. Detailní požadavky na poskytování služeb v rámci záruky jsou uvedeny v příloze č. 3 závazného vzoru smlouvy, který tvoří přílohu č. 3 zadávací dokumentace.
		5. Instalací a zprovozněním se rozumí instalace hardware do stávajících rack skříní, zapojení všech síťových rozhraní, zapojení do elektrické sítě, nastavení BIOSu a BMC dle pokynů zadavatele a spuštění hardware a ověření bezchybného chodu všech komponent. Pokud bude dodavatel preferovat prověření výkonových požadavků na vlastní instalaci linuxových strojů, musí být součástí instalace a zprovoznění také instalace příslušného počtu klientských stanic.
	2. **Detailní technické požadavky na uzly clusterů - každý výpočetní uzel clusteru musí splňovat tyto podmínky:**
		1. Provedení do standardního 19" racku, velikost uzlu musí být maximálně 1U. Servery budou umístěny do stávajících racků, racky nejsou součástí dodávky dle této veřejné zakázky.
		2. Sdílení komponent mezi více uzly clusteru v jedné lokalitě je povoleno, ale výpadek žádné sdílené komponenty nesmí ovlivnit více jak 4 uzly.
		3. V případě sdílení aktivních komponent, tj. komponent obsahujících ventilátory nebo pracující s napětím elektrické sítě (např. zdroje a jejich přívody), musí být všechny tyto komponenty redundantní. Redundance komponent v jednotlivých uzlech není požadována.
		4. Každý uzel clusteru (výpočetní jednotka se samostatnou pamětí, chipsetem, procesory, diskem atd.) musí mít alespoň jeden procesor se sdílenou pamětí v architektuře x86\_64 a minimálně 64 fyzických jader celkem.
		5. Minimální požadovaný výkon celého uzlu měřený nástrojem SPECfp2017 ve variantě FP, rate, baseline je 500 bodů. Zároveň výkon v tomto benchmarku, přepočtený na jedno jádro CPU, tj. výkon celého uzlu vydělený počtem fyzických jader v uzlu (počítají se pouze fyzická jádra, nikoli technologie hyperthreading), musí být alespoň 7.
		6. Výkon lze prokázat předložením oficiálního výsledku z webu [www.spec.org](http://www.spec.org/) dosaženého na ekvivalentním stroji (typ a počet procesorů totožný s dodanými procesory; počet a frekvence paměťových DIMMů a organizace paměti totožné s dodanou pamětí, velikost paměti se může lišit) nebo výsledkem spuštění benchmarku na uzlu nakonfigurovaném dle uvedené technické specifikace.
		7. Nabízený procesor i další komponenty (motherboard, BIOS) musí podporovat virtualizaci, včetně virtualizace I/O (např. VT-d v terminologii firmy Intel, AMD-Vi v terminologii firmy AMD).
		8. Operační paměť každého uzlu alespoň 8GB ECC na jedno fyzické jádro procesoru. Rychlost pamětí nesmí být horší než rychlost paměti použité ve SPEC benchmarku v odst. 1.4.5. Na všech použitých paměťových kanálech musí být stejná skladba DIMMů.
		9. Každý uzel clusteru musí mít přístup k lokálnímu NVMe prostoru, určenému pro operační systém a lokální dočasná data, vše realizováno alespoň dvěma totožnými SSD NVMe disky s celkovou kapacitou alespoň 100 GB na každé fyzické jádro procesoru. Součet rychlostí lineárního čtení/zápisu všech disků musí být alespoň 10/6 GB/s, součet I/O výkonu všech disků musí být alespoň 2000k/200k IOPS pro náhodné čtení/zápis. Součet výdrže disků musí být alespoň 13 PB zápisů (13000TBW), případná reklamace disku nesmí být zamítnuta kvůli jeho opotřebení zápisy, pokud dodavatel neprokáže, že byl překročen alikvotní podíl z 13PB odpovídající podílu jednotlivého disku na celkové kapacitě všech disků.
		10. Všechny disky musí podporovat vyčtení provozních stavů a statistik pomocí standardu SMART.
		11. Každý uzel clusteru musí mít alespoň dvě 10 Gb ethernet rozhraní typu BaseT s RJ45 konektorem. Ethernetová rozhraní musí podporovat UDP TNL offload v linuxovém ovladači (příznaky NETIF\_F\_GSO\_UDP\_TUNNEL a NETIF\_F\_GSO\_UDP\_TUNNEL\_CSUM) a bootování přes PXE.
		12. V lokalitě MU požadujeme v každém uzlu alespoň jeden volný PCIe 4.0 x16 slot pro low-profile (167.65mm x 68.90mm) pro případné budoucí osazení infinibandové karty.
		13. Každý uzel clusteru musí umožňovat přístup ke konzoli (klávesnice + monitor) jak lokálně tak po síti (síťový KVM nebo BMC).
		14. Každý uzel clusteru musí podporovat bootování z externího zařízení, a to jak lokálně (boot z USB – CD-ROM, flash disk, harddisk), tak po síti (síťový KVM nebo BMC).
		15. Základní deska musí umožňovat změnu pořadí bootovacích zařízení.
		16. Základní deska musí obsahovat management controller (BMC) kompatibilní se specifikací IPMI 2.0 nebo vyšší. BMC musí umět monitorovat minimálně funkčnost ventilátorů a zdroje, teplotu CPU a základní desky; dále musí BMC poskytovat základní vzdálený power management (vypnout, zapnout, reset). Požadujeme možnost změny bootovacího zařízení vzdáleně pomocí BMC nebo KVM.
		17. Grafické rozhraní KVM (BMC nebo externího) musí fungovat ve webovém prohlížeči s podporou HTML5.
		18. Vzdálený přístup k BMC musí být realizovatelný buď po sdíleném síťovém kabelu (tj. každý uzel je připojený do sítě pouze dvěma 10Gb ethernetovými kabely, jeden z nich sdílí BMC provoz) nebo samostatným kabelem. V případě použití samostatného kabelu pro BMC je dodavatel povinen dodat do každé lokality (MU i ÚMG) i switche s minimálním počtem portů odpovídajícím počtu uzlů v dané lokalitě plus 3 a rychlostí portů alespoň 1Gb/s.

Při použití více switchů v jedné lokalitě musí být switche stohovatelné a při instalaci nakonfigurované dodavatelem do jediného stohu (stohem se rozumí sestava propojených switchů, která se při komunikaci navenek, včetně funkcí vzdálené správy, chová jako jeden switch s potřebným počtem portů). V lokalitě Brno připouštíme kabeláž mezi racky pouze ve formě stackovacího kabelu.

* 1. **Požadavky na síťové komponenty**
		1. Součástí nabídky musí být veškerá propojovací kabeláž pro připojení všech výpočetních uzlů clusteru do sítových switchů, tj. 2x10Gb kabely pro běžný provoz a je-li použita varianta samostatného kabelu (viz 1.4.18) tak i 1Gb kabely pro BMC. Všechny kabely kalkulujte na míru, při instalaci je žádoucí, aby nevznikaly zbytečné kabelové smyčky.
1. **2. Další požadavky:**
	1. Veškerá zařízení by mělo být možno koupit a používat ke stanoveným účelům bez jakéhokoliv přídavného software (externího, nepočítá se firmware, jehož trvalá licence je součástí dodávky zařízení). Pokud je programové vybavení nutnou součástí nabízeného plnění (například SW pro vzdálenou správu), musí být jasně specifikovány důvody a cena za oprávnění k užívání takového software musí být zahrnuta do nabídkové ceny (oprávnění bude uděleno na dobu neurčitou, resp. na dobu trvání majetkových práv autora).
	2. Všechny komponenty, které jsou touto technickou specifikací požadovány, musí být použitelné v prostředí operačního systému Linux (zejména, ale nikoliv výhradně 64bit Debian), tj. musí být podporovány distribučním nebo originálním jádrem nebo s využitím externích ovladačů dostupných ve zdrojovém kódu.
	3. V nabídce musí být uvedena celková maximální spotřeba sestavy v každé lokalitě (maximální spotřeba odpovídá spotřebě při plném zatížení všech komponent, tedy serverů).
	4. V nabídce musí být uvedena velikost jednotlivých instalací v každé lokalitě v počtu U ve standardním 19'' racku.
	5. Všechna zařízení instalovaná v racku musí mít výstup teplého vzduchu směřovaný dozadu racku.
	6. Dodavatel musí nakonfigurovat EFI/BIOS všech serverů jednotně dle požadavku zadavatele tak, aby stroj bylo možné nabootovat z PXE a následně z pevného disku.
	7. U zařízení umožňujících vzdálenou správu předá dodavatel zadavateli seznam s umístěním těchto zařízení v racku (pozice v racku, MAC adresa) a provede nastavení BMC dle pokynů zadavatele (jednotné heslo, IP adresa z dodaného rozsahu, VLAN…) tak, aby bylo možné stroje ovládat na dálku.
	8. Vybraný dodavatel bude povinen na výzvu zadavatele prokázat původ dodaného plnění a další skutečnosti, viz odst. 5.10. závazného vzoru smlouvy, který tvoří přílohu č. 3 zadávací dokumentace.
	9. Veškeré vyvázání kabeláže bude provedeno způsobem umožňujícím opakované beznástrojové rozebrání a znovupoužití.
	10. Veškeré kabely musí být opatřeny na obou koncích shodným, v rámci instalace jednoznačným, označením, aby bylo možné identifikovat odkud kam vedou bez nutnosti procházet celou kabelovou trasu.
	11. Požadujeme instalaci do stávajících stojanů (racků) v místě instalace tak, aby v nich zůstaly zbylé volné pozice v co největších souvislých blocích.
	12. Pro instalaci jsou k dispozici:
		1. **v lokalitě MU** dva racky celkové výšky 42U a vnější šířky 80 cm a hloubky 120cm (rozvor vertikálních instalačních lišt je možné upravit), zády přímo proti sobě přes teplou uličku šířky 120cm. V každém racku jsou trvale obsazeny tři nejvyšší pozice, k dispozici je 39U volných pozic. Napájení je redundantní, každá ze dvou větví má celkem 3\*32A 230V (komponenty, u kterých jsou požadovány redundantní napájecí zdroje musí být schopny trvalého běhu jen na jednu větev a celá sestava nesmí překročit proudový limit ani při zapnutí po neřízeném výpadku napájení) rozdělené na 28 zásuvek typu CEE7/5 (česká). Dostupná kapacita chlazení celkem max. 27kW na rack, příkon celé sestavy ji nesmí přesahovat ani při trvalém maximálním využití. Kabely realizující požadované propoje mezi prvky v různých stojanech musí být vedeny přes kabelovou lávku ve výšce 70 cm nad střechami racků a vodorovné vzdálenosti 40 cm od prostupů ve střechách racků. Potřebná délka případného stackovacího kabelu (1.4.18) může v závislosti na poloze portů propojovaných switchů přesáhnout 5m.
		2. **v lokalitě ÚMG** dva racky APC celkové výšky 42U a vnější šířky 75 cm a hloubky 120cm (rozvod vertikálních instalačních lišt je možné upravit) v jedné řadě, oddělené pouze in-row chladícími jednotkami nebo distribučními panely o šířce 30 cm. Proudění chladného vzduchu je ve směru z přední do zadní části racku. V každém racku jsou trvale obsazeny dvě nejvyšší pozice, k dispozici je 40U volných pozic. Napájení je redundantní, každý rack je v zadní části vybaven čtyřmi vertikálními PDU 32A/230V (komponenty, u kterých jsou požadovány redundantní napájecí zdroje musí být schopny trvalého běhu jen na jednu větev a celá sestava nesmí překročit proudový limit ani při zapnutí po neřízeném výpadku napájení) rozdělené na 21 zásuvek IEC 60320 C13 (celkem je tedy k dispozici 21\*4 = 84 zásuvek na rack) . Dostupná kapacita chlazení celkem max. 40kW (tj. 20kW na rack), příkon celé sestavy ji nesmí přesahovat ani při trvalém maximálním využití. Nejsou k dispozici průchodky přímo mezi racky, kabely realizující požadované propoje mezi prvky v různých stojanech musí být vedeny přes střechy racků, délka kabelů musí umožňovat vedení po kabelové lávce ve výšce 30 cm nad střechami racků a vodorovné vzdálenosti 20 cm od zadních lišt racků.