**Załącznik nr 2**

SZCZEGÓŁOWY OPIS PRZEDMIOTU ZAMÓWIENIA – SPECYFIKACJA TECHNICZNA

**Zakup i dostawa serwerów obliczeniowych wyposażonych w kartę GPU**

**Przedmiotem zamówienia jest dostawa 2 takich samych serwerów przeznaczonych do obliczeń związanych ze sztuczną inteligencją, w szczególności do uruchamiania dużych modeli językowych (LLM). Każdy z serwerów musi spełniać poniższe parametry:**

1. Serwer musi posiadać obudowę przystosowaną do montażu w szafie rack 19”, o wysokości maksymalnej 3U i głębokości maksymalnej 120cm.
2. Obudowa musi umożliwiać instalację przynajmniej 8 dysków SAS/SATA/NVMe
3. Serwer musi być wyposażony w minimum 4 redundantne zasilacze hot-swap o mocy nie mniejszej niż 2600 W każdy, z certyfikatem minimum Titanium, Hot-Swap.
4. Serwer musi umożliwiać zasilanie napięciem przemiennym 230 V / 50 Hz.
5. Serwer musi być wyposażony w co najmniej 2 porty sieciowe 1 Gbps oraz co najmniej 2 porty sieciowe optyczne o przepustowości minimum 10 Gbps każdy, z możliwością pracy w środowisku Ethernet.
6. Karty sieciowe nie powinny zajmować żadnego z dostępnych slotów PCIe.
7. Serwer musi być wyposażony w co najmniej dwa dyski M.2 SSD NVMe o pojemności minimum 960 GB każdy, podłączone do dedykowanego sprzętowego kontrolera RAID i przeznaczone do uruchamiania systemu operacyjnego, z możliwością konfiguracji w RAID 1. Dodatkowo serwer powinien posiadać cztery dyski NVMe o pojemności 1,92 TB, podłączone bezpośrednio do backplane.
8. Wymagany jest wewnętrzny slot na kartę Micro SD.
9. Serwer musi być wyposażony w pamięć RAM o pojemności co najmniej 1 TB, o częstotliwości 6400MHz, z możliwością dalszej rozbudowy do 3TB, Zabezpieczenie pamięci ECC, SDDC, Patrol/Demand Scrubbing.
10. Serwer musi być wyposażony w zintegrowaną kartę graficzną z minimum 16MB pamięci osiągająca rozdzielczość 1920x1200 przy 60 Hz, karta powinna być wyposażona w 1 port VGA na tylnym panelu.
11. Serwer musi być wyposażony w dwa procesory klasy serwerowej, posiadający minimum 16 rdzeni każdy. Minimalna częstotliwość pracy pojedynczego procesora to 3.0GHz. Wynik wydajności procesora zainstalowanego w oferowanym serwerze nie powinien być niższy niż 870 punktów base w teście SPECrate 2017 Integer opublikowanym przez SPEC.org (www.spec.org) dla konfiguracji dwuprocesorowej. Test przeprowadzony przez producenta serwera musi być zamieszczony na stronie spec.org.
12. Serwer musi być wyposażony w jedną kartę przyspieszające obliczenia równoległe (GPU), wyposażoną w minimum 140 GB pamięci HBM, obsługujące formaty obliczeń zmiennoprzecinkowych FP32 o wydajności nie mniejszej niż 60 teraFLOPS, oraz FP64 o wydajności nie mniejszej niż 30 teraFLOPS.
13. Format karty GPU: DW (double-wide) PCIe, Gen5, karta dwuslotowa, chłodzona powietrzem
14. Minimalna wydajność Tensor Core (dla różnych precyzji):
    1. TF32 Tensor Core co najmniej 835 teraFLOPS
    2. BFLOAT16 / FP16 Tensor Core co najmniej 1,671 teraFLOPS
    3. FP8 Tensor Core co najmniej 3,341 teraFLOPS
15. Tryb współdzielenia GPU (MIG) do 7 instancji (Multi-Instance GPUs), każda o ~16.5 GB pamięci
16. Serwer powinien umożliwiać instalację 4 kart przyśpieszających obliczenia równoległe.
17. Serwer powinien być wyposażony w min. 3 sloty PCIe x16, oraz 4 sloty dedykowane do instalacji kart GPU podwójnej szerokości, oraz dedykowany slot dla kart sieciowych (OCP).
18. Dedykowane sloty GPU powinny umożliwiać przesyłanie danych bezpośrednio do innych kart PCIe z pominięciem CPU.
19. Wraz z serwerem powinno zostać dostarczone dodatkowe oprogramowanie zarządzające umożliwiające:
    1. zarzadzanie infrastruktura serwerówi storage bez udziału dedykowanego agenta
    2. przedstawianie graficznej reprezentacji zarządzanych urządzeń
    3. możliwość skalowania do minimum 1000 urządzeń
    4. dostępność konsoli monitorującej na urządzeniach przenośnych ze wsparciem dla systemu Android oraz iOS, aplikacja musi umożliwiać włączenie, wyłączenie, oraz restart urządzenia, musi również mieć możliwość aktywowania diody lokacyjnej na urządzeniu
    5. konsola zarzadzania oparta o HTML 5
    6. monitorowanie oraz tracking zużycia energii przez monitorowane urządzenie
    7. możliwość podnoszenia wersji oprogramowania dla komponentów zarządzanych serwerów w oparciu o repozytorium lokalne jak i zdalne dostepne na stronie producenta oferowanego rozwiązania
20. Jeżeli oprogramowanie wymaga dodatkowej licencji, powinno ono zostać dostarczone wraz z serwerem oraz okres trwania licencji powinien pokrywać się z okresem gwarancji.
21. Serwer musi być objęty gwarancją obejmującą:
    1. dostęp do aktualnych wersji oprogramowania sprzętowego,
    2. realizację zgłoszeń serwisowych i gwarancyjnych w trybie 24×7,
    3. w godzinach 9:00-17:00 wymagana jest obsługa zgłoszeń przez wsparcie techniczne producenta w języku polskim.
    4. wsparcie techniczne świadczone telefonicznie, mailowo i za pośrednictwem portalu www,
    5. dostarczenie sprawnego urządzenia w terminie maksymalnie 4 dni roboczych od momentu zgłoszenia awarii.
    6. uszkodzone dyski pozostają u Zamawiającego