

Specyfikacja wymagań technicznych dla klastra obliczeniowego 50 Tflops

Wymagania podstawowe dla klastra.

- teoretyczna moc obliczeniowa sumaryczna klastra nie mniejsza niż 50 Tflops, gdzie teoretyczną moc obliczeniową R_{peak} należy obliczyć ze wzoru:

$$R_{peak} = R_{proc} * liczba procesorów$$

- maksymalny pobór mocy 8 kW na szafę,
- maksymalny całkowity pobór mocy 220 kW,
- wszystkie nody klastra (węzły) muszą być jednego typu oraz muszą posiadać identyczne procesory,

Konfiguracja serwerów obliczeniowych, z których zbudowany jest klaster.

Przez serwer obliczeniowy Zamawiający rozumie niezależną jednostkę, posiadającą odrębny zasilacz, złożoną z jednego lub więcej **nodów**, zamkniętą w odrębnej obudowie typu RACK, przy czym:

- wysokość każdego serwera (poza serwerem dostępowym, opisanym poniżej) nie może być większa niż 1U (do montowania w szafie teleinformatycznej typu RACK),
- każdy serwer musi być wyposażony w zestaw do zamontowania serwera na wysuwanych szynach w szafie teleinformatycznej 19,”

Nody klastra

Przez nod klastra (węzeł, element obliczeniowy) Zamawiający rozumie pojedynczą płytę główną wraz z procesorami, pamięcią, dyskiem twardym oraz ewentualnymi kartami rozszerzeń, wymagającą osobnego podłączenia sieci Gigabit Ethernet i InfiniBand,

Każdy nod musi posiadać:

- procesory, pamięć operacyjną, pamięć masową według poniższych specyfikacji,
- 2 interfejsy sieciowe Gigabit Ethernet o prędkości 10/100/1000 Mbit/s (medium: miedź, RJ45). Dozwolona jest konfiguracja w postaci pojedynczej karty dwuportowej,
- minimum jeden interfejs InfiniBand o przepustowości min. 20 Gb/s (w technologii DDR, medium miedziane) umożliwiający podpięcie węzła do przełącznika pracującego w technologii InfiniBand,
- możliwość podłączenia zewnętrznego napędu DVD-ROM przez interfejs USB lub zamontowany wewnętrzny napęd DVD-ROM,
- 1 wolny slot umożliwiający zainstalowanie karty rozszerzeń w standardzie *PCI-Express x8*,
- możliwość podpięcia każdego z nodów niezależnie do konsoli KVM, poprzez port VGA zintegrowany na płycie głównej węzła,
- możliwość bootowania systemu operacyjnego przez sieć (*DHCP + TFTP*),
- możliwość bootowania systemu operacyjnego z urządzenia USB typu pen-drive,

- możliwość zdalnego załączania i wyłączenia nodu oraz monitorowanie parametrów pracy (temperatura, napięcia, prędkości wentylatorów),
- możliwość dostępu do konsoli szeregowej przez sieć Ethernet (*Serial-over-LAN*),
- pełne wsparcie systemu Linux 64-bit, jądro minimum 2.6, dla węzła i zainstalowanych w nim kart rozszerzeń,

Wszystkie serwery oraz nody (węzły) obliczeniowe muszą pochodzić od jednego producenta, być nowe, jednego typu oraz muszą posiadać identyczne procesory.

Procesory w nodzie

- minimum 4-rdzeniowe,
- w architekturze 64-bitowej typu *x86-64/EM64T* lub *IA64*,
- maksymalnie po 2 procesory na pojedynczą płytę główną,
- teoretyczna moc obliczeniowa pojedynczego procesora nie mniejsza niż 35 GFlops. Teoretyczną moc obliczeniową procesora należy obliczyć ze wzoru:

$$R_{proc} = C * I * F$$

R_{proc} - moc obliczeniowa w GFlops

C - liczba rdzeni procesora

I - liczba instrukcji zmiennoprzecinkowych typu dodawanie i mnożenie (w podwójnej precyzji) wykonywanych przez pojedynczy rdzeń procesora w czasie jednego cyklu zegarowego (np. dla procesora Intel Itanium2 dual-core (Montecito) i Intel Xeon Woodcrest (seria 5100) n wynosi 4, natomiast dla procesorów AMD Opteron n wynosi 3, dla pozostałych modeli Intel Xeon n wynosi 2)

F - częstotliwość zegara w GHz

Pamięć operacyjna w nodzie

- minimum 4 GB pamięci RAM na każdy procesor 4-rdzeniowy,
- typ pamięci: najwyższy typ pamięci (np. DDR2) oraz maksymalne taktowanie dostępne wg specyfikacji producenta dla danego zestawu „płyta-procesor”, w modułach po min. 2GB lub więcej,
- możliwość rozbudowy pamięci do minimum 32 GB na każdy węzeł,

Pamięć masowa w nodzie

- na każdą płytę główną co najmniej jeden dysk *SATA*, *SCSI* lub *SAS* o pojemności minimum 160 GB i prędkości obrotowej minimum 7200 RPM

Konfiguracja serwera dostępowego

Jeden serwer dostępowy o następującej konfiguracji:

- 2 procesory identyczne jak w węzłach obliczeniowych,
- 16 GB pamięci operacyjnej RAM z możliwością rozbudowy do min. 32 GB,
- obudowa typu rack o wysokości 2U,
- 4 dyski twarde *SATA* o prędkości obrotowej co najmniej 7200 RPM i sumarycznej pojemności nie mniejszej niż 2400 GB,
- 2 dyski twarde *SCSI* lub *SAS* po minimum 140 GB pojemności dyskowej każdy, o prędkości obrotowej 10000 RPM lub większej,

- kontroler RAID, SCSI lub SAS, w zależności od dostarczonych dysków, umożliwiający podłączenie wszystkich zaoferowanych dysków SCSI lub SAS w macierz dyskową, obsługujący tryby RAID przynajmniej 0, 1 i 10 (dopuszcza się kontroler zintegrowany na płycie głównej lub w formie karty rozszerzeń),
- kontroler *RAID SATA* umożliwiający podłączenie wszystkich zaoferowanych dysków *SATA* w macierz dyskową, obsługujący tryby *RAID* przynajmniej 0, 1 i 5 (dopuszcza się kontroler zintegrowany na płycie głównej lub w formie karty rozszerzeń),
- 4 porty Gigabit Ethernet (dopuszczalne jest zastosowanie jednej dwuportowej karty rozszerzeń),
- minimum 1 port InfiniBand w technologii min. 20 Gb/s DDR (dopuszczalne jest zastosowanie karty rozszerzeń),
- możliwość zdalnego załączania i wyłączania nodu oraz monitorowanie parametrów pracy (temperatura, napięcia, prędkości wentylatorów),
- możliwość dostępu do konsoli szeregowej przez sieć Ethernet (*Serial-over-LAN*),
- wbudowany napęd DVD-ROM z możliwością zapisu,
- redundantne zasilacze.

Sieć dla aplikacji obliczeniowych

Sieć typu InfiniBand:

- przełącznik lub zespół przełączników pracujących w technologii InfiniBand o przepustowości minimum 20 Gb/s na port, umożliwiające podłączenie wszystkich nodów,
- konfiguracja przełącznika powinna zapewniać nieblokującą architekturę *Full-CBB* dla całego klastra, w standardzie DDR,
- urządzenia muszą być przystosowane do montażu w szafach rack 19",
- Zamawiający wymaga aby Dostawca przedstawił w ofercie szczegółowy opis proponowanego rozwiązania wraz z topologią połączeń.

Sieć do zarządzania i składowania danych

Sieć w technologii Gigabit Ethernet o przepustowości 1 Gb/s na port Gigabit Ethernet w każdym nodzie:

- oparta na przełącznikach 48-portowych,
- przynajmniej jeden przełącznik 48-portowy na każdą szafę do połączenia nodów oraz jeden dodatkowy przełącznik do połączenia struktury drzewa,
- przełączniki muszą się posiadać następujące cechy:
 - szybkość przełączania magistrali: minimum 39 Mpps,
 - obsługa protokołu SNMP w wersji 3,
 - możliwość tworzenia VLANów w standardzie *IEEE 802.1q*,
 - obsługa *Rapid Spanning Tree Protocol IEEE 802.1w*,
 - obsługa *Multiple Spanning Tree Protocol IEEE 802.1s*,
 - możliwość zarządzania przez interfejs WWW oraz z poziomu wiersza poleceń (CLI) z dostępem przez SSH i/lub TELNET,
 - możliwość montażu w szafie typu rack 19",
 - wysokość 1U.

System operacyjny

- zainstalowany system operacyjny *Debian GNU/Linux* w wersji 64-bitowej, odpowiedniej do architektury procesorów, w najnowszej stabilnej wersji
- pełna współpraca sprzętu z systemem operacyjnym Linux
- wsparcie producenta sprzętu dla systemu operacyjnego Linux
- monitoring i zarządzanie działające pod Linuxem

Oprogramowanie i biblioteki

- kompilatory języków C, C++ oraz Fortran 90/95 zalecane przez producenta dostarczanych procesorów (np. *PathScale Compiler Suite* dla procesorów *AMD* lub *Intel Compilers* dla procesorów *Intel*) wraz z licencjami
- implementacja biblioteki BLAS, poziom 1, 2 i 3 zoptymalizowana dla dostarczanych procesorów, dostarczana przez producenta procesora (np. *Intel Math Kernel Library*, *AMD Core Math Library*) wraz z licencjami
- debugger aplikacji umożliwiający sprawdzanie aplikacji MPI (np. *Intel Trace Analyzer and Collector 7.0*)

Zarządzanie

- przynajmniej jedna 8-portowa konsola KVM z ekranem LCD na każde 120 serwerów, przystosowana do montażu w szafie typu rack 19"
- dodatkowo zestaw złożony z monitora LCD, klawiatury, myszy

Oprogramowanie do zarządzania i monitorowania klastra umożliwiający:

- zdalne włączenie, wyłączenie i restart węzła
- autoryzację użytkowników
- administrację logami
- administrację ustawieniami sieciowymi
- pełen zdalny dostęp konsolowy umożliwiający: pracę na węźle bez pośrednictwa systemu operacyjnego, przekierowanie standardowej konsoli (ttyX) w przypadku systemu Linux
- zdalne monitorowanie podstawowych parametrów węzła
- wykonanie ww. operacji (oprócz dostępu konsolowego) na wielu węzłach jednocześnie z zastosowaniem skryptów lub innego oprogramowania (powinno to być możliwe także z uwzględnieniem autoryzacji użytkowników)
- oprogramowanie powinno posiadać interfejs konsolowy i uruchamiać się na oferowanej platformie z wykorzystaniem systemu Debian Linux

W przypadku gdy wykonawca napisze odpowiednie skrypty lub oprogramowanie we własnym zakresie, udostępni je Nabywcy wraz ze źródłami. Nabywca zastrzega sobie możliwość wprowadzania zmian w ww. oprogramowaniu na swoją odpowiedzialność, bez konsultacji z wykonawcą i bez ponoszenia dodatkowych kosztów z tym związanych.

Kable, urządzenia porządkujące okablowanie, dodatkowe akcesoria.

- odpowiednia liczba paneli zasilających umożliwiających zdalne monitorowanie parametrów zasilania i umożliwiających zdalne sterowanie załączaniem/wyłączaniem serwerów,
- odpowiednia ilość patchcordów cat5e lub lepszych z wtykami RJ45 (montowane fabrycznie, długość i kolory do ustalenia przed dostawą) umożliwiająca połączenie serwerów z przełącznikiem Gigabit Ethernet,
- odpowiednia ilość opasek na rzepy do spinania kabli,
- odpowiednia ilość trwałych etykiet z tworzywa sztucznego do oznaczenia kabli sieciowych ethernetowych na obu końcach, zawierające: numer szafy, pozycję węzła w szafie,
- odpowiednia ilość kabli w standardzie InfiniBand 4X, umożliwiająca połączenie nodów z przełącznikami InfiniBand i zapewnienie w/w topologii sieci obliczeniowej
- korytka podtrzymujące kable prowadzone pomiędzy szafami i przełącznikami w obrębie szaf według potrzeb,
- zewnętrzny napęd DVD-ROM podłączany przez interfejs USB.

Szafy.

Szafy teleinformatyczne 19” o sumarycznej pojemności odpowiedniej dla całości sprzętu objętego dostawą (serwery, przełączniki, akcesoria szaf, itd.) z zachowaniem wymogu maksymalnego poboru mocy 8 kW na szafę wg następującej specyfikacji:

- model ZPAS SZB SE serii szaf serwerowych 19” w wykonaniu drzwi przednie blaszane perforowane, lub równoważny. Jako model równoważny Zamawiający przyjmuje szafę o identycznych parametrach technicznych oraz wizualnych (materiały i standard wykonania, włączając grubości i typ zastosowanych blach oraz sposób wykonania i wyglądu, m.in. perforacji) z szafami ZPAS SZB SE posiadanymi już przez Zamawiającego. Zamawiający pragnie zachować wizualną zgodność szaf ze względów estetycznych.
- wymiary: 600mm x 1000mm x 36U (szerokość/głębokość/wysokość użytkowa),
- wysuwany cokół zabezpieczający przed przewróceniem się szafy,
- wymagane zdejmowane drzwi przednie (do każdej szafy) oraz osłony boczne z blachy perforowanej (do skrajnych szaf w szeregach), dach pełny,
- perforacja drzwi i osłon musi być wykonana na całości powierzchni (oprócz ramy konstrukcyjnej) przy zachowaniu jak największych otworów dla maksymalnej cyrkulacji powietrza,
- wymagana możliwość demontażu szafy na czas transportu,
- wymagana możliwość trwałego złączenia szaf bokami w celu zapewnienia stabilności i estetyki zespołu szaf, przy zamknięciu osłonami bocznymi szeregu,
- wszystkie szafy jednego typu i w jednym kolorze: czarny

Szczegółowy zakres prac będących przedmiotem zamówienia.

- rozładunek urządzeń,
- fizyczny montaż urządzeń w miejscu wskazanym przez Zamawiającego,
- wykonanie niezbędnych połączeń elektrycznych i logicznych w uzgodnieniu z Zamawiającym. Zamawiający wymaga określenia, z co najmniej dwutygodniowym wyprzedzeniem przed planowanym dniem dostawy, typu przyłącza elektrycznego szafy rack, oraz wymagań co do prowadzenia okablowania pomiędzy szafami,
- instalacja i konfiguracja na każdym z dostarczonych nodów systemu operacyjnego Debian

- GNU/Linux. Konfiguracja systemu obejmuje w swym zakresie wszelkie czynności niezbędne do uruchomienia oprogramowania wymienionego w kolejnych sześciu punktach,
- instalacja systemu kolejkowego PBS (Torque 2.1.8) w konfiguracji 3 kolejek: 15 minut, 24 godziny i 3 dni,
 - instalacja oprogramowania MAUI wersja 3.2.6,
 - instalacja i konfiguracja oprogramowania MPICH wersja 1.2.7 oraz wersja 2,
 - instalacja i konfiguracja oprogramowania MVAPICH2 wersja 0.9.8,
 - instalacja i konfiguracja oprogramowania OpenMPI wersja 1.2.2,
 - instalacja i konfiguracja oprogramowania do zarządzania infrastrukturą serwerów,
 - instalacja, uruchomienie i przeprowadzenie testu wydajności klastra za pomocą programu HPL,
 - przeszkolenie pracowników Zamawiającego,
 - opracowanie i dostarczenie dokumentacji powykonawczej,

Szkolenie.

Zamawiający wymaga przeszkolenia minimum 3 pracowników w zakresie administracji dostarczonym klastrem na poziomie podstawowym, w lokalizacji ustalonej z Wykonawcą, Miejsce, termin oraz szczegółowa tematyka szkolenia zostaną ustalone przez Strony w terminie do 4 tygodni od daty podpisania protokołu odbioru sprzętu,

Dokumentacja powykonawcza.

Zamawiający wymaga dostarczenia, po zakończeniu całości projektu, dokumentacji powykonawczej systemu zawierającej co najmniej:

- dokumentację techniczną dostarczonych urządzeń,
- szczegółowy opis konfiguracji dostarczonego oprogramowania oraz urządzeń,
- procedury eksploatacyjne oraz awaryjne.