

Szacowanie zasobów obliczeniowych

- [Podstawowe informacje](#)
- [Jakie zasoby szacować?](#)
- [Szacowanie czasu obliczeniowego](#)
 - [Klasy maszyn w PLGrid](#)
 - [Testowe zadania](#)
 - [Wybór klasy węzłów podczas zlecenia zadania](#)
 - [Obliczanie czasu potrzebnego w grantcie](#)

Podstawowe informacje

- Oszacowanie potrzebnych zasobów dobrze jest przeprowadzić **przed zawnioskowaniem o grant**.
- Wynik oszacowania pomoże użytkownikowi w wyborze grantu ([testowego](#) lub [właściwego](#)) oraz w uzupełnieniu wniosku o grant właściwy.
- Podczas szacowania ilości potrzebnych zasobów użytkownik nabywa wiedzę o udostępnianych w PLGrid zasobach (maszynach, dyskach, maszynach wirtualnych).
- Dobre oszacowanie potrzebnych w grantcie zasobów:
 - wpływa pozytywnie na ocenę z [rozliczenia](#) grantu,
 - sprawia, że użytkownik nie potrzebuje [renegocjować](#) grantu,
 - pomaga administratorom zasobów w ich adekwatnym dystrybuowaniu,
 - może wpłynąć na szybsze wykonywanie się zadań użytkownika.

Jakie zasoby szacować?

Użytkownik - zanimawnioskując o grant właściwy - może oszacować, ile i jakie zasoby spodziewa się zużyć podczas badań w ramach grantu.

Aby w pełni wykorzystać możliwości Infrastruktury PLGrid, dobrze jest **zastanowić się nad każdą** metryką w grantcie.

Sądząc po wykorzystaniu zasobów w grantcie testowym lub na dotychczas używanym sprzęcie, użytkownik może z grubsza powiedzieć, ile potrzebuje pamięci dyskowej, jak bardzo jest w stanie zrównoleglić zadania obliczeniowe, ile RAM-u będą one zajmować itp. Najważniejsze jednak to oszacować czas obliczeniowy, potrzebny na wykonanie się pojedynczego i wszystkich zadań.

Szacowanie czasu obliczeniowego

[Czas obliczeniowy](#) w grantcie mierzymy w godzinach znormalizowanych. Różni się one od godzin zegarowych specjalnym przelicznikiem - [współczynnikiem normalizacji](#).

Aby dobrze oszacować czas obliczeniowy, potrzebny w grantcie, należy wziąć pod uwagę:

- heterogeniczność Infrastruktury PLGrid - oferujemy różne klasy maszyn o różnej mocy na różnych klastrach PLGrid,
- [ofertę zasobów PLGrid](#),
- czas wykonania pojedynczego zadania,
- równoległość wykonywanych zadań.

Klasy maszyn w PLGrid

W Infrastrukturze PLGrid oferujemy zasoby klastrów z polskich ośrodków: **Cyfronet, ICM, PCSS, TASK i WCSS**. Klastry te składają się (z grubsza) z serwerów o różnej mocy obliczeniowej. To oznacza, że jeden klastrowy może oferować maszyny obliczeniowe o różnych procesorach (ich różnej mocy, a nawet różnej architekturze). To oznacza też, że klastry między ośrodkami różnią się w zakresie mocy i architektury procesora.

Zanim Użytkownik wybierze do obliczeń klastrowy lub klastry, może **zapoznać się z ofertą PLGrid**, gdzie wyszczególniono każdy klastrowy, jego klasy węzłów oraz m.in. ich procesory, dostępny RAM oraz tagi węzłów w systemie kolejkowym.

[Klasa węzłów](#) klastrowy to zbiór węzłów obliczeniowych o tych samych parametrach (np. o tym samym modelu procesora, jednakowej RAM per węzeł i tym samym tagu w systemie kolejkowym).



Wyboru klastrów w oparciu o ofertę należy dokonać w oparciu o zapotrzebowanie, np.:

- Jeśli zależy nam na konkretnym modelu procesora (gdyż mamy już przystosowane do niego oprogramowanie, skompilowane dla danej architektury), można w ofercie wyszukać ośrodki oferujące dany model procesora i jego architekturę (informacje o niej zawarto w tagu systemu kolejkowego).
- Jeśli nasze zadania wymagają sporego zapotrzebowania na RAM, sprawdźmy w ofercie parametr RAM per węzeł.
- Jeśli obliczenia wymagają użycia kart graficznych, sprawdźmy w ofercie, jakie klastry posiadają klasy węzłów oferujące GPGPU.

Gdy zdecydujemy się na obliczenia na konkretnym klastrze lub klastrach, pora na prosty test zadań na wybranych zasobach.

Testowe zadania

Jeśli Użytkownik wybrał już ośrodki obliczeniowe i klasy maszyn w ramach klastrów w ośrodkach, proponujemy teraz:

1. Zawnioskować o [grant testowy](#) (lub skorzystać z aktywnego grantu testowego).
2. W ramach grantu testowego wykonać skrypt (skrypty), który będzie potem uruchamiany w ramach obliczeń w [grancie właściwym](#) (na większą skalę). Skrypt należy wykonać na [wybranych wcześniej klasach węzłów](#) w ośrodkach.
3. Sprawdzić wykorzystanie zasobów przez zlecane zadanie. Wykorzystanie pojawi się na stronie [Accounting PLGrid](#) oraz na wybranych klastrach za pomocą [odpowiednich komend](#).
4. Ewentualnie powtórzyć zlecenie zadań - od punktu 2. W celu uzyskania dokładniejszych danych zadanie można zlecać kilkakrotnie na jedną klasę, a następnie uśrednić czas trwania zadania na wybranej klasie.
5. Oszacować na podstawie testu liczbę godzin obliczeniowych potrzebną w grancie właściwym.
 - a. Czas wykonania zadania należy znormalizować.
 - b. Otrzymany czas znormalizowany wykonania się zadania na danej klasie węzłów należy przemnożyć przez szacunkową liczbę zadań potrzebnych do otrzymania wyników.
 - c. Jeśli zadania zaplanowano do wykonania na kilku różnych klasach węzłów, należy przemnożyć szacunkową liczbę zadań na danej klasie przez znormalizowany czas wykonywania się zadania na tej klasie węzłów, a następnie zsumować czasy wykonania się wszystkich zadań na wszystkich pożądanym klasach węzłów.
 - d. Otrzymany czas znormalizowany wykonania się wszystkich zadań należy przemnożyć przez liczbę rdzeni - jeśli zadania będą zlecane wielordzeniowo.

Wybór klasy węzłów podczas zlecania zadania

Aby zlecane zadanie wykonało się na pożądanym klasie węzłów, należy "powiadomić" o tym system kolejkowy. W tym celu należy:

- wpisać informację o klasie węzłów w skrypcie zlecającym zadanie lub
- podać informację o klasie węzłów w komendzie zlecającej zadanie.

Każda klasa węzłów posiada oznaczenie (tag) w systemie kolejkowym. Oznaczenie to jest widoczne w ostatniej kolumnie [oferty PLGrid](#). Np. klasa węzłów Prometheusa w Cyfronecie o procesorach Intel Xeon E5-2680 v3 posiada oznaczenie: "haswell_2500mhz".

Przykład skryptu z informacją o zleceniu zadania na węzeł danej klasy:

Przykładowy skrypt zadania

```
#!/bin/bash
##ID grantu, w ramach którego zlecane bd zadania testowe
#PBS -A plguzytkownik2016a
##liczba wykorzystywanych wzów (nodes) i rdzeni (ppn) przez zlecane zadanie - tu: 1 rdze klasy
haswell_2600mhz w PCSS (klaster Inula)
#PBS -l nodes=1:ppn=1:haswell_2600mhz
#PBS -l walltime=00:15:00
#PBS -q plgrid-testing
#PBS -m abe -M plguzytkownik@domena.pl

cd $PBS_O_WORKDIR

echo "start szacowania -" $(date)
#w tym miejscu nalezy uruchomi zadanie, wywoujc jego skrypt lub adujc modu oprogramowania i wywoujc program
z podanym inputem zadania
echo "koniec szacowania -" $(date)

exit 0;
```

Szczegółowe informacje o zlecaniu zadań do systemu kolejkowego znajdują się w sekcji: [Uruchamianie zadań bezpośrednio przez system kolejkowy \(SSH\)](#).

Czas wykonywania się zadania, otrzymany w skrypcie powyżej, jest czasem rzeczywistym, który upłynie w trakcie wykonywania zadania, należy go więc znormalizować [współczynnikiem normalizacji](#) odpowiednim dla klasy węzłów, na której zadanie się wykonywało.

Obliczanie czasu potrzebnego w grantcie

Po wykonaniu obliczeń na wszystkich wybranych klasach maszyn należy przeprowadzić oszacowanie.

Aby uzyskać szacunek czasu obliczeniowego potrzebnego w grantcie należy wykonać obliczenia dla przybliżonej liczby zadań, które zostaną uruchomione w ramach grantu, a następnie należy:

- rzeczywisty czas trwania zadania przemnożyć przez współczynnik normalizacji dla danej klasy węzłów,
- przemnożyć otrzymany czas znormalizowany zadania przez liczbę rdzeni, które zadanie będzie wykorzystywać,
- a jeśli zadania w grantcie będą uruchamiać się na kilku różnych klasach węzłów, należy następnie podsumować ich czasy znormalizowane w zależności od klasy węzłów, na której się uruchomią.

Na przykład:

Obliczenia będą prowadzone na Prometheusie (Cyfronet) i na Bemie (WCSS). Jedno zadanie będzie wykonywać się ok. 30 min. W ramach obliczeń planujemy zlecić 1000 zadań na 24 rdzenie każde.

Klasa maszyn	Współczynnik normalizacji a	Czas rzeczywisty wykonania zadania b	Czas znormalizowany c	liczba zadań d	wielordzeniowość zadań e	Znormalizowany czas zadań
Prometheus - haswell_2500mhz	3,84	30 min	$a * b = 3,84 * 30 \text{ min} = 115,2 \text{ min} (1,92 \text{ h})$	1000	24	$c * d * e = 1,92 \text{ h} * 1000 * 24 = 46\ 080 \text{ h}$
Bem - haswell_2300mhz	3,74	35 min	$a * b = 3,74 * 35 \text{ min} = 131,25 \text{ min} (2,19 \text{ h})$	1000	24	$c * d * e = 2,19 \text{ h} * 1000 * 24 = 52\ 560 \text{ h}$
Oszacowany czas obliczeń w grantcie właściwym:						98 640 h

W grantcie potrzeba ok. **98 640 h**.



Podczas wnioskowania o grant należy wpisać czas obliczeniowy według powyższych wyliczeń (lub większy, z niewielkim zapasem - np. 100 000 h). We wniosku o grant, w polu "Uzasadnienie wykorzystania zasobów" można opisać swoje wyliczenia.