



Dziedzinowo zorientowane  
usługi i zasoby infrastruktury  
PL-Grid dla wspomagania  
Polskiej Nauki w Europejskiej  
Przestrzeni Badawczej

# Implementation of Computing Resources Normalization in PL-Grid Infrastructure

**R. Różańska, A. Pułapa, A. Golik, M. Radecki**

ACK Cyfronet

KU KDM Zakopane, 13.03.2014



**Czas obliczeniowy** - miara zużycia zasobów przez zadania użytkownika. Jest on obliczany jako iloczyn czasu trwania obliczeń oraz liczby wykorzystanych przez niego rdzeni.

Zliczanie czasu obliczeniowego:

- do 1.01.2014 czas obliczeniowy wyrażany był w **godzinach zegarowych** (czas rzeczywisty) – brak uwzględnienia mocy maszyny, na którą zadanie zostało zlecone
- od 1.01.2014 czas obliczeniowy wyrażany jest w **godzinach znormalizowanych** (czas znormalizowany) – uwzględniana jest wydajność używanych procesorów w rozliczaniu zleconych zadań, dzięki czemu możliwe jest **sprawiedliwe rozliczenie użytkownika** - podobne zadania, których czas działania różnił się z powodu różnic w wydajności węzłów na których zostały uruchomione powinny być potraktowane jako równoważne

Dwa zadania zlecone na węzły o **różnej mocy**, wykonane w **takim samym czasie**, będą miały przypisane **różne wykorzystanie**. Zadanie wykonane na **węźle szybszym** zużyje **więcej godzin znormalizowanych**, niż zadanie wykonane na węźle wolniejszym. Różnica będzie proporcjonalna do różnicy mocy węzłów.

Czas zadania Klasa Maszyn	Rzeczywisty (wsp. norm. 1)	Znormalizowany
Intel(R) Xeon(R) CPU L5640 @ 2.27GHz (wsp. norm. 0,97)	100 h	97 h znorm.
Intel(R) Xeon(R) CPU X5670 @ 2.93GHz (wsp. norm. 1,31)	100 h	131 h znorm.

Aby uzyskać znormalizowany czas zleconego zadania należy wykonać następujące kroki:

## 1. wyznaczyć współczynniki normalizacji dla klas maszyn dostępnych w Infrastrukturze

- obliczanie **mocy pojedynczego rdzenia** (na podstawie przeprowadzonych benchmarków):

$$\text{moc rdzenia (GFlops)} = \frac{\text{moc węzła (GFlops)}}{\text{liczba rdzeni}}$$

- obliczanie **współczynnika normalizacji dla klasy**:

$$\text{współczynnik normalizacji} = \frac{\text{moc rdzenia (GFlops)}}{\text{średnia ważona moc rdzenia w całej Infrastrukturze (GFlops)}}$$

2. pomnożyć czas obliczeniowy zadania (w godzinach zegarowych) przez współczynnik normalizacji charakterystyczny dla węzła (wyliczony w poprzednim kroku), na którym wykonywało się zadanie

- Obliczenie ***znormalizowanego czasu*** trwania zadania:

*$znormalizowany\ walltime = czas\ rzeczywisty\ trwania\ zadania \times\ współczynnik\ normalizacji$*

## Zakładka „Normalizacja” w Portalu PLGrid:

Accounting **Normalizacja**

Normalizacja

### Wyniki benchmarków

CYFRONET-LCG2				
Nazwa węzła	Hash	Liczba rdzeni w węźle	Liczba węzłów	Współczynnik normalizacji
Intel(R) Xeon(R) CPU L5420 @ 2.50GHz	169d6c25e238264763ed867f1d6fdbb7	8	164	1.01
AMD Opteron(TM) Processor 6276	4e2a27b16b11063260f4ea380fb10c62	64	101	0.84
Intel(R) Xeon(R) CPU X5670 @ 2.93GHz	613c459b21c510792ab30f2dc10e92e7	12	22	1.31
Intel(R) Xeon(R) CPU L5640 @ 2.27GHz	9585ee73a1e6cbdefe380aaedd159229	12	242	0.97
Intel(R) Xeon(R) CPU X5650 @ 2.67GHz	c4cee7e20ed23a8c77f8582c424f7ab6	12	340	1.15
Intel(R) Xeon(R) CPU E5645 @ 2.40GHz	fca9f662cd10de9dc82fcdc384caaca1	12	276	1.04

ICM-HYDRA				
Nazwa węzła	Hash	Liczba rdzeni w węźle	Liczba węzłów	Współczynnik normalizacji
Six-Core AMD Opteron(tm) Processor 2435	4aa11b72094ce9c79db239c66721e7ad	12	95	0.99
AMD Opteron(tm) Processor 6174	74f7f5fb28b963a5b5b3bf973f92caf9	48	29	0.88
Intel(R) Xeon(R) CPU X5660 @ 2.80GHz	9c3991f6047a7e2d13c6e41f419a8c30	12	120	1.21
AMD Opteron(tm) Processor 6134	a92bf88311333833e82dc1ca0548d7eb	16	20	0.87

- zlecając zadanie obliczeniowe, Użytkownik może **wskazać klasę maszyn**, na którą chce zlecić zadanie: podanie parametru **-l** wraz z charakterystyką żądanych zasobów podczas zlecenia zadania

## Przykład zlecenia zadania z parametrem -l

```
qsub -l nodes=1:ppn=8:mhz2500
```

Wyspecyfikowane powyżej parametry to:

- liczba węzłów (nodes),
- liczba rdzeni alokowanych dla zadania (ppn - nie może być większa niż liczba rdzeni na węzeł danej klasy),
- oznaczenie klasy węzłów charakteryzujące zasoby (tutaj: mhz).

Powyższa specyfikacja oznacza, że zadanie uruchomi się na **jednym** węźle klastra Zeus, na **8 rdzeniach** **węzła**, na klasie maszyn z procesorami **Intel Xeon L5420** (co specyfikuje własność węzła: mhz2500).

Opis zasobów na stronie:

[http://www.plgrid.pl/oferta/zasoby\\_obliczeniowe/opis\\_zasobow/HPC](http://www.plgrid.pl/oferta/zasoby_obliczeniowe/opis_zasobow/HPC)

# Klasy maszyn – parametry przy zlecaniu zadań



## Prezentacja zasobów obliczeniowych Infrastruktury

### HPC

#### Systemy obliczeniowe

Opisy dla poszczególnych pól w kolumnie "Uwagi" wyświetlają się w postaci *tooltip* po najechaniu kursorem myszy. Opisy te są również dostępne pod tabelą.

Ośrodek	Nazwa systemu	Liczba węzłów danego typu	Liczba procesorów w każdym węźle	Liczba rdzeni każdego procesora	Liczba rdzeni w każdym węźle	Model procesora	GHz	Wielkość pamięci całego węzła [GB]	Wielkość pamięci w przel. na każdy rdzeń [GB]	Oznaczenie w systemie kolejkowym (node properties)	Uwagi
ACK CYFRONET AGH	Zeus	256	2	4	8	Intel Xeon L5420	2,5	16	2	mhz2500, mem16gb, core8	
		256	2	6	12	Intel Xeon L5640	2,26	16	1,33	mhz2266, mem16gb, core12, ib_ql, ib_all	
		342	2	6	12	Intel Xeon X5650	2,66	24	2	mhz2666, mem24gb, core12, ib_mlx, ib_all, infiniband	
		234	2	6	12	Intel Xeon E5645	2,4	24	2	mhz2400, mem24gb, core12, ib_mlx, ib_all, infiniband	
	6	2	6	12	Intel Xeon E5645	2,4	192	16	mhz2400, mem192, core12, ib_mlx, ib_all, infiniband, bigmem	C1	
	Zeus vSMP	64	2	6	12	Intel Xeon X5650	2,66	96	8	mhz2666, vsmp	C1





- **podając klasę maszyny**, na której zadanie ma się wykonać, czas rzeczywisty wykonania zadania **przemnażany jest przez współczynnik normalizacji** danej klasy
- w przypadku **niepodania klasy** maszyny, zadanie **uruchomi się na przypadkowym węźle**.
- w przypadku, gdy węzeł, na którym uruchomi się zadanie, **nie jest przypisany do żadnej klasy** maszyn (np. jest nowy i jeszcze nie został sklasyfikowany), zadanie zostanie potraktowane jako wykonujące się na **najwolniejszym węźle na klastrze** (obliczony czas znormalizowany będzie korzystny dla użytkownika)
- jeśli żadna maszyna na klastrze **nie posiada współczynnika normalizującego**, wówczas za **współczynnik dla całego klastra przyjmuje się 1** (czas znormalizowany jest równy czasowi zegarowemu)

# Prezentacja wyniku normalizacji (Accounting)



- Od dnia 1 stycznia 2014 wykorzystanie zasobów, znajdujące się w systemie accountingowym PL-Grid (Accounting) jest domyślnie **wyświetlane wykorzystanie zasobów w godzinach znormalizowanych**. Wykorzystanie w godzinach zegarowych dostępne jest w oddzielnej kolumnie (domyślnie ukrytej, możliwej do rozwinięcia) oraz po najechaniu myszką na komórkę z wykorzystaniem znormalizowanym.
- Normalizacja obejmuje jedynie wykorzystanie zasobów **po 1 stycznia 2014**, co znaczy, że zadania, które rozpoczęły się wcześniej, nie podlegają normalizacji (dla nich współczynnik normalizacji wynosi 1).

grant2013	2013-05	0	0	0	0	0
opsscripts	2013-07	0				
plgrid	2013-07	71				
undefined	2013-07	72				
undefined	2013-08	5				
plgabaldyga2013a	2013-08	0	Unnormalized value: 1			
testsystgrant4	2013-10	1	1	1	10	10
		1495				



# Szacowanie potrzebnych zasobów obliczeniowych



- pozwala na bardziej precyzyjne wnioskowanie o **grant właściwy** w PLGrid
- uwzględnia **klasy maszyn**, na które planuje się zlecać zadania związane z grantem
- uwzględnia **równoległość zadań** w grantcie
- przykładowy **skrypt zlecający zadanie** pomocny w szacowaniu zasobów
- wnioskowana liczba godzin **bliższa wykorzystaniu zasobów w grantcie**

Chętnych pragnących zapoznać się z instrukcją szacowania zapraszamy na stronę

<https://docs.cyfronet.pl/pages/viewpage.action?pageId=10063747>



# DZIĘKUJEMY!