



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

# Wykorzystanie infrastruktury chmurowej i platformy GridSpace w obliczeniach naukowych

Daniel Harężlak, Eryk Ciepiela, Jan Meizner  
i Marian Bubak

ACK CYFRONET AGH

konferencja i3, Poznań, 16 kwietnia 2013



- Planowanie wirtualnych eksperymentów
- GridSpace
  - Struktura eksperymentów
  - Model publikowania
- Wymagania dla infrastruktur chmurowych
- Struktura chmurowego eksperymentu
  - Zarządzanie wirtualnymi maszynami
  - Kontrola parametrów środowiska uruchamiania usług
- Podsumowanie

# Planowanie wirtualnego eksperymentu – przypadki użycia GridSpace



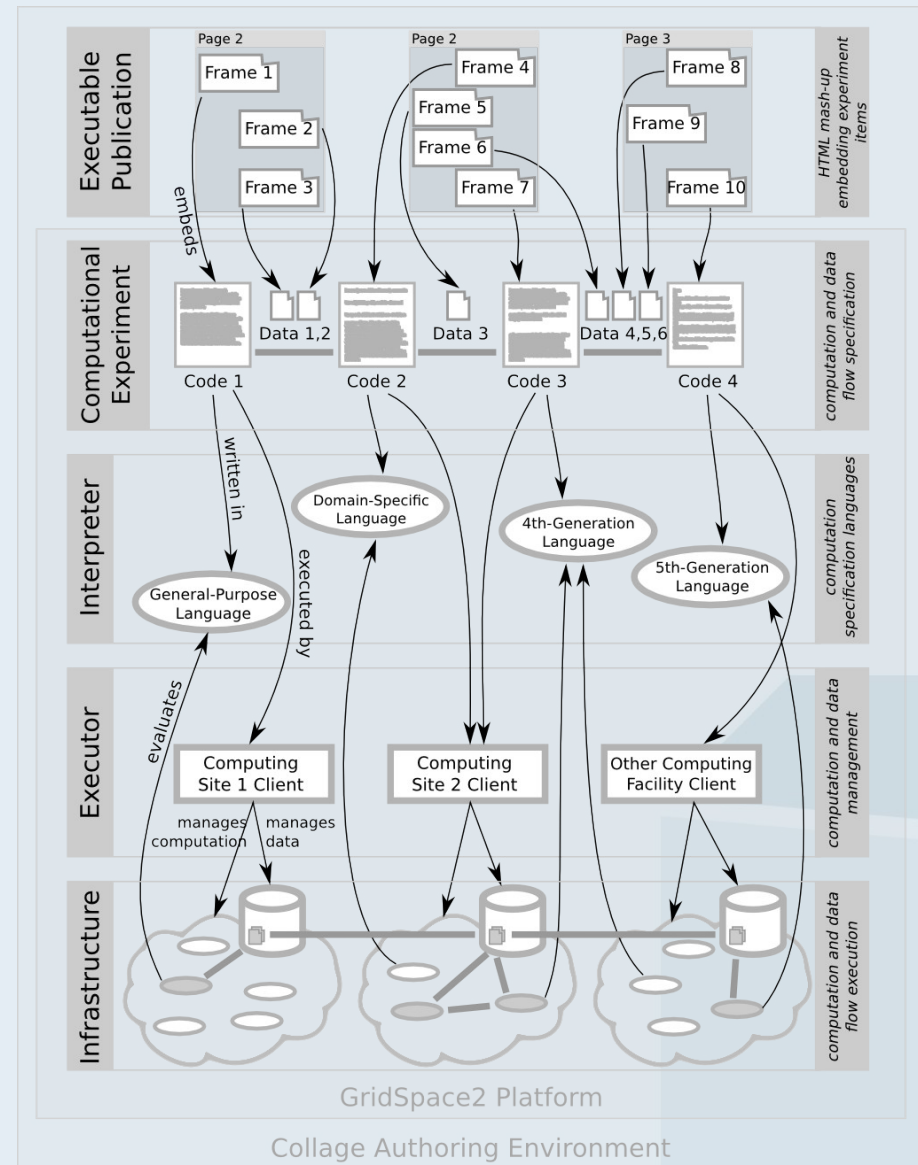
- Praca kolaboratywna i wspólna analiza rezultatów pośrednich
  - Potrzeba dzielenia się kodem i danymi z różnymi prawami dostępu
- Dynamiczna infrastruktura obliczeniowa i zmieniające się środowisko uruchomienia
- Wykorzystanie rozproszonej infrastruktury obliczeniowej
  - Zadania lokalne, zadania gridowe, dedykowane maszyny obliczeniowe, wykorzystanie różnych *middleware*
- Użycie gotowych pakietów oprogramowania, autorskich kodów i bibliotek lub ich kombinacji
- Testowanie różnych zasobów obliczeniowych i szybkości ich działania (ten sam kod, różne środowiska uruchomienia)
- Integracja ze współczesnymi systemami wymiany informacji opartych o środowiska webowe (np. strony wiki, blogi, itp.)



# GridSpace – Struktura eksperymentów



- Publikowalne w warstwie *web* elementy wirtualnego eksperymentu
- Wirtualny eksperyment składający się z wykonywalnych kodów oraz rezultatów pośrednich i końcowych
- Konfigurowane administracyjnie pakiety aplikacji i interpretery
- Rozszerzalny system (*plugins*) dostępu do różnych stosów *middleware*
- Zapewniony dostęp do różnych infrastruktur obliczeniowych i pamięci masowej za pomocą jednej zintegrowanej platformy

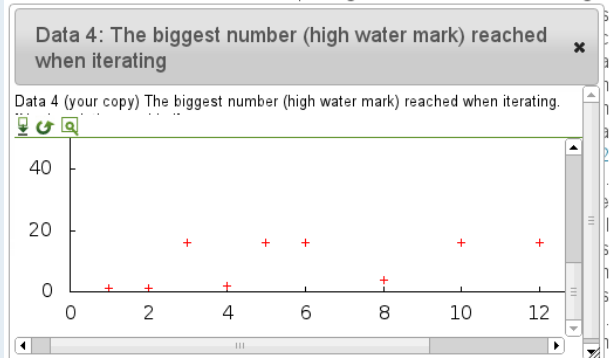


## The Collatz Conjecture

eciepiela, DOI: 10.0000/1358511059290

The experiment was released by eciepiela on Fri Jan 18 13:10:59 CET 2013 in the private scope. No, below is not an article. It's only generated text with injected labels that navigate to particular experiment items to show you how in-text links work.

Euismod in pellentesque massa placerat dui ultricies lacus. Penatibus et magnis dis parturient. Ipsum consequat nisi vel pretium lectus. Interdum velit euismod in pellentesque. Nascetur ridiculus mus mauris vitae ultricies leo integer malesuada nunc. Sapien nec sagittis aliquam malesuada. Donec ultrices tincidunt arcu non sodales neque sodales. Sed velit dignissim sodales ut eu. [Data 1](#) Ac tortor vitae purus faucibus ornare. Mattis enim ut tellus elementum sagittis vitae et. Vitae semper quis lectus nulla at. Vitae purus faucibus ornare suspendisse sed nisi lacus sed viverra. Amet tellus cras adipiscing enim eu. [Code 1](#) Tellus integer



non nisi est sit amet. Forbitor massa tur neque airquam vestibulum morbi blandit. [Data 5](#) Non consectetur a erat nam at lectus urna duis. Diam volutpat commodo sed egestas egestas fringilla phasellus faucibus scelerisque. Magna eget est lorem ipsum dolor sit.

Data 1: Arguments

Code 1: Computing sequences for given arguments

```
Code 1 (your copy) Computing sequences for
Source: Ruby 1.8.7 Output
step+=1
end
results_raw.puts("#(argument)\t#\t#")
```

Data 2: Raw results to be visualized afterwards

Data 2 (original) Raw results to be visualized

1	0	1	1
2	1	1	1
3	7	16	1
4	2	2	1
5	5	16	1
6	8	16	1
7	10	52	10

Code 2: Generating plots

Data 3: Number of iterations

Data 4: The biggest number (high water mark) reached when iterating

Data 5: Value in last iteration

Save experiment Reset

- Publikacja wykonywalnych kodów i danych referencyjnych
- Konfiguracja infrastruktury obliczeniowej zapisana w eksperymencie
- Możliwość ponownego uruchomienia i weryfikacji opublikowanych treści
- Dostęp do elementów eksperymentu za pomocą przeglądarki

Collage Authoring Workbench

You are logged in to collage-exphost.elsevier.com as eciepiela

collage-exphost.elsevier.com: collatz/collatz

The Collatz Conjecture

Computing sequences for Ruby 1.8.7 with collage-exphost.elsevier.com

Generating plots GnuPlot 4.2.6 with collage-exphost.elsevier.com

```
set output 'collatz/hwm.png'
plot 'collatz/results_raw.txt' using 1:3 title "The biggest number (high water mark) reached when iterating"

set output 'collatz/last.png'
plot 'collatz/results_raw.txt' using 1:4 title "Value in last iteration"
```

1 input data defined for this code. 3 output data defined for this code.

Output

```
gnuplot> set output 'collatz/hwm.png'
gnuplot> plot 'collatz/results_raw.txt' using 1:3 title "The biggest number (high water mark) reached when iterating"
```

Filter by file name

Path: eciepiela/collatz/

Releases

# Wymagania dla infrastruktury chmurowych – przypadek usługi porównującej białka



- Znana ilość wymaganej pamięci operacyjnej w zależności od wielkości problemu
- Usługa udostępniona za pomocą kontenera Tomcat (Java)
  - Trudna instalacja na dostępnych infrastrukturach obliczeniowych
- Długość wykonania względnie krótka (dziesiątki sekund)
  - Czas niewspółmierny z czasem oczekiwania w lokalnych lub gridowych kolejkach
- Udostępnienie usługi na żądanie w trakcie wykonania wirtualnego eksperymentu



# Struktura chmurowego eksperymentu – Zarządzanie wirtualnymi maszynami



- Przygotowanie obrazu maszyny wirtualnej
  - Instalacja wymaganego oprogramowania
- Zapisanie obrazu i udostępnienie w platformie chmurowej PL-Grid PLUS
- Konfiguracja i uruchomienie z wirtualnego eksperymentu w kontekście konkretnego użytkownika
  - Żądana jest maszyna o wymaganej ilości pamięci operacyjnej
  - Konfigurowane są przekierowania portów w celu udostępnienia maszyny na zewnątrz
- Wyłączenie instancji wirtualnych maszyn
  - Zapewnione jest poprawne sprzątnięcie i oszczędność zasobów obliczeniowych





# Struktura chmurowego eksperymentu – Parametry środowiska uruchomienia



- Dynamiczna generacja deskryptora opisującego parametry instancji maszyny wirtualnej z poziomu eksperymentu
  - Żądanie określonej ilości pamięci operacyjnej
- Przygotowanie przekierowań portów z poziomu eksperymentu
  - Udostępnienie usługi na zewnątrz z prywatnej sieci platformy chmurowej
- Przykład deskryptora:

```
CONTEXT=[  
  SSH_KEY="..." ]  
CPU="0.5"  
DISK=[  
  IMAGE_ID="7" ]  
GRAPHICS=[  
  TYPE="vnc" ]  
MEMORY="1024"  
NAME="AUTO7-plgharezlak-1364909661"  
NIC=[  
  NETWORK_ID="0" ]  
OS=[  
  ARCH="x86_64" ]  
TEMPLATE_ID="77"
```





## ■ Najważniejsze atuty platformy GridSpace

- Skonfigurowane różnorodne infrastruktury obliczeniowe (lokalne kolejki, zadania gridowe, AHE, QCG)
- Możliwość łatwego rozszerzania poprzez dodatkowe pluginy
- Model publikowania rozszerzony o wykonywalny kod i dane
- Łatwa integracja z infrastrukturą chmurową PL-Grid PLUS

## ■ Planowane prace

- Usprawniona integracja z systemem uwierzytelniania użytkowników
- Dedykowany *executor* chmurowy komunikujący się bezpośrednio z API platformy chmurowej

## ■ Więcej informacji na następujących stronach

<http://dice.cyfronet.pl>

<https://gs2.plgrid.pl>

- E. Ciepiela, D. Harezlak, J. Kocot, T. Bartynski, M. Kasztelnik, P. Nowakowski, T. Gubała, M. Malawski, M. Bubak: **Exploratory Programming in the Virtual Laboratory**. In: Proceedings of the International Multiconference on Computer Science and Information Technology (2010)
- P. Nowakowski, E. Ciepiela, D. Hareźlak, J. Kocot, M. Kasztelnik, T. Bartyński, J. Meizner, G. Dyk, M. Malawski: **The Collage Authoring Environment**. In: Proceedings of the International Conference on Computational Science (2011)
- E. Ciepiela, P. Nowakowski, J. Kocot, D. Harezlak, T. Gubała, J. Meizner, M. Kasztelnik, T. Bartynski, M. Malawski, M. Bubak: **Managing Entire Lifecycles of e-Science Applications in the GridSpace2 Virtual Laboratory - From Motivation through Idea to Operable Web-Accessible Environment Built on a Top of PL-Grid e-Infrastructure**. In: M. Bubak, T. Szepieniec, K. Wiatr (Eds) Building a National Distributed e-Infrastructure - PL-Grid - Scientific and Technical Achievements, Springer (2012)
- D. Harezlak, M. Kasztelnik, E. Ciepiela, M. Bubak: **Scripting Language Extensions Offered by the GridSpace Experiment Platform**. In: M. Bubak, T. Szepieniec, K. Wiatr (Eds) Building a National Distributed e-Infrastructure - PL-Grid - Scientific and Technical Achievements, Springer (2012)
- J. Meizner, E. Ciepiela, P. Nowakowski, J. Kocot, M. Malawski, M. Bubak: **Flexible and Extendable Mechanism Enabling Secure Access to e-Infrastructures and Storage of Confidential Data for the GridSpace2 Virtual Laboratory**. In: M. Bubak, T. Szepieniec, K. Wiatr (Eds) Building a National Distributed e-Infrastructure - PL-Grid - Scientific and Technical Achievements, Springer (2012)
- E. Ciepiela, P. Nowakowski, D. Hareźlak, M. Kasztelnik, G. Dyk, J. Meizner, M. Bubak: **The Collage Authoring Environment: A Platform for Executable Publications**, abstract for Cracow Grid Workshop (2012)
- E. Ciepiela, D. Hareźlak, M. Kasztelnik, J. Meizner, G. Dyk, P. Nowakowski, M. Bubak: **The Collage Authoring Environment: from proof-of-concept prototype to an exploitable service. Development and deployment of a web-oriented platform for executable scientific publications**, accepted to International Conference on Computational Science (2013)
- P. Pierzchała: **Multiscale Applications in the GridSpace Virtual Laboratory**, Master of Science Thesis supervised by Katarzyna Rycerz; AGH University of Science and Technology (2012)
- B. Bodziechowski, E. Ciepiela, M. Bubak: **Assessment of Software Quality with Static Source Code Analysis: GridSpace2 Case Study**, abstract for Cracow Grid Workshop (2012)