

Platforma HyperFlow

Uwaga!

Ze względu na wycofanie usługi Cloud 2.0, platforma HyperFlow jest dostępna w formie właściwej aplikacji dostępnej w repozytorium: <https://github.com/hyperflow-wms> która może zostać uruchomiona w ramach Cloudu 3.0 lub dowolnego z klastrów obliczeniowych. Repozytorium kodu zawiera obecnie dostępne skrypty, wspomagające automatyzację uruchomienia platformy HyperFlow. Poniższy przykładowy opis uruchomienia platformy w ramach Zeusa nie jest już preferowaną metodą korzystania z HyperFlow.

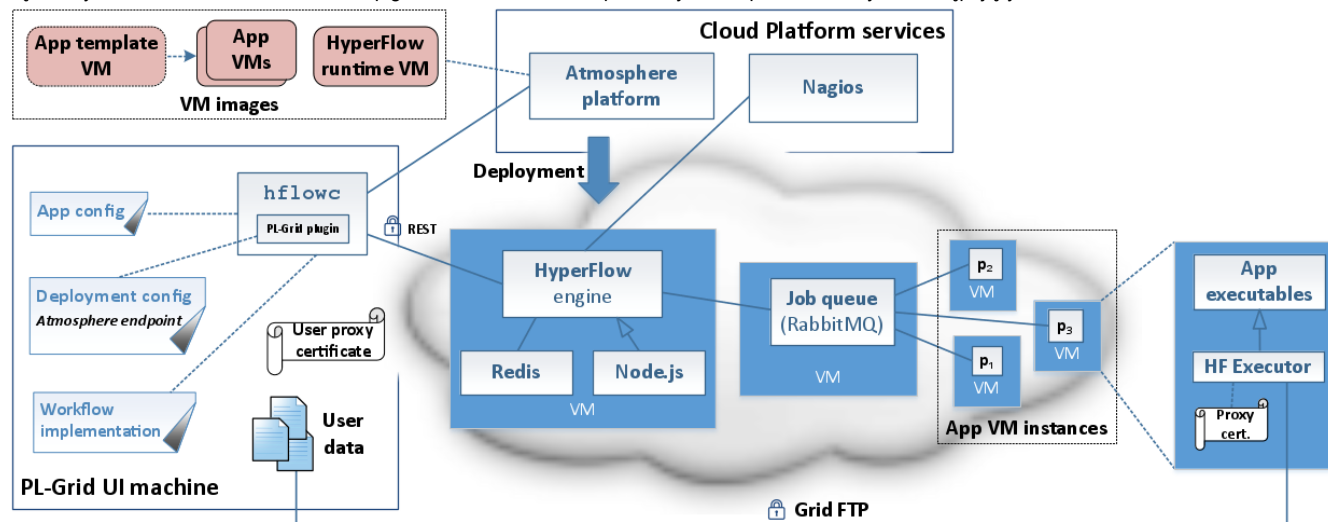
Krótki opis usługi

Platforma HyperFlow jest usługą przeznaczoną zarówno dla zaawansowanych deweloperów jak i naukowców dziedzinowych, pracujących nad rozwojem aplikacji reprezentowanych przez workflowy. Gotowe workflowy korzystające z platformy mogą być wykorzystywane przez zwykłego użytkownika infrastruktury PLGrid.

Usługa oferuje "platformę", czyli bazę narzędziową służącą do rozwoju aplikacji typu workflow. Developer powinien rozpoznać dostarczone narzędzia, dostosować je do swoich potrzeb i zdefiniować aplikację w postaci workflowu. Poszczególne narzędzia oferują funkcjonalności typowe dla swojej roli, czyli np. automatyczny transfer plików wejścia wyjścia, zlecenie zadań na zdalne maszyny, delegacja uprawnień itd. Baza zawiera gotowe elementy systemu, które można wykorzystać podczas budowania swojej aplikacji. Gotowe elementy to między innymi:

- Zestaw obrazów maszyn wirtualnych, dostępnych na cloudzie:
 - GD workflow main - maszyna zawierająca rdzeń wykonania workflowu w postaci:
 - silnik wykonawczy HyperFlow
 - serwer kolejek Rabbitmq
 - GD workflow worker - maszyna będąca wykonawcą zadań, zawiera:
 - aplikację "executor", która odpowiada za pobranie zadań z kolejki i wykonanie ich
- Klient hflowc, który służy do automatyzacji tworzenia środowiska wykonawczego w cloudzie.

Korzystając z w/w komponentów możliwe jest skonstruowanie aplikacji, która będzie zdefiniowana poprzez workflow a jej środowisko wykonawcze będzie dynamicznie tworzone w cloudzie plgrid. Schemat działania platformy został przedstawiony na następującym obrazku:



Aktywowanie usługi

Aby aktywować usługę "Platforma HyperFlow", postępuj zgodnie z instrukcjami zawartymi w [Katalogu Aplikacji i Usług](#) po uprzednim wyszukaniu usługi. Przed zaaplikowaniem o tą usługę konieczne jest posiadanie certyfikatu gridowego, wystawionego przez SimpleCA i następujących usług:

- Dostęp do platformy Cloud Computing - Cloud 2.0
- Dostęp do klastra ZEUS.

Pierwsze kroki

Poniższy przykład prowadzi użytkownika poprzez podstawowy scenariusz użycia platformy. Przykład pozwala na skorzystanie ze wszystkich komponentów.

We wszystkich przykładach należy zastąpić login swoim loginem, wygenerowane URLe będą się różnić od przykładowych.

Stworzenie definicji workflowu:

Należy utworzyć w swoim katalogu domowym plik 'workflow.json' o następującej zawartości:

```
{
  "name": "hello_task",
  "processes": [ {
    "name": "process",
    "type": "dataflow",
    "function": "amqpCommand",
    "config": {
      "executor": {
        "executable": "/bin/cat",
        "args": "input.txt > output.txt",
        "options": {
          "prefix": "gsiftp://zeus.cyfronet.pl/people/plglogin"
        }
      }
    }
  },
  "ins": [ 0 ],
  "outs": [ 1 ]
}],
  "signals": [ {
    "name": "input.txt",
    "data": "1"
  }, {
    "name": "output.txt"
  } ],
  "ins": [ 0 ],
  "outs": [ 1 ]
}
```

lub pobrać gotowy plik:

[workflow.json](#)

Przygotowanie pliku wejściowego:

```
[plglogin@zeus ~]$ echo "input" > input.txt
```

Setup środowiska:

```
[plglogin@zeus ~]$ module load plgrid/tools/hflowclient

[plglogin@zeus ~]$ grid-proxy-init
(...)
[plglogin@zeus ~]$ hflowc setup
Appliance set hfworkflow created successfully!
wfMain created successfully!
waiting for wfMain to start.....wfMain is active!
Found port mapping 1032->443, hfmain endpoint: http://149.156.11.4:1032
starting wfworker_0
Workers created
```

Notujemy "hfmain endpoint", po zakończeniu wykonania polecenia należy odczekać jeszcze 1-2 minuty aby wszystkie usługi się podniosły. W międzyczasie możemy obserwować tworzenie się maszyn w interfejsie webowym cloudu, dostępnym pod adresem: <https://cloud.plgrid.pl> w zakładce "Workflows". Na tym etapie, w cloudzie powstało środowisko uruchomieniowe.

Uruchomienie workflowu:

```
[plglogin@zeus ~]$ hflowc runwf http://149.156.11.4:1032 workflow.json
workflow started: http://149.156.11.4:1032/apps/1
```

W tym momencie workflow zostanie przesłany do maszyny zawierającej silnik HyperFlow i uruchomiony. Wynikiem działania workflowu będą informacje znajdujące się w pliku wyjściowym.

Sprawdzenie wyników:

W następujący sposób sprawdzamy zawartość pliku wynikowego, który został stworzony na "workerze" w trakcie wykonania workflowu a następnie skopiowany jako "output" do katalogu zdefiniowanego w workflowie.

```
[plglogin@zeus ~]$ cat output.txt
hello world
```



Usunięcie środowiska uruchomieniowego workflow:

```
[plgpawlik@zeus ~]$ hflowc teardown
Removing appliance set: hfworkflow
Done.
```

Zaawansowane użycie

Zaawansowane użycie sprowadza się do zmodyfikowania obrazów, tak aby zawierały narzędzia potrzebne do wykonania zadań z workflowu (konkretne aplikacje np. montage), zapisanie ich i modyfikacja konfiguracji klienta, tak aby nowo tworzone środowiska były budowane z uwzględnieniem własnych obrazów.

Nazapki

Podczas korzystania z klienta otrzymuję komunikat podobny do:

```
[Error: error getting appliance sets: 400, Bad Request]
```

Należy się upewnić, że korzystamy z proxy od SimpleCA stworzonego za pomocą komendy `grid-proxy-init` lub `voms-proxy-init -rfc` lub proxy pobranego z PLGridowego OpenId (format: RFC 3820 compliant impersonation proxy).

Gdzie szukać dalszych informacji?

Szczegółowa dokumentacja dla elementów systemu jest dostępna w następujących lokalizacjach:

- <https://github.com/dice-cyfronet/hyperflow> (opis silnika wykonawczego HyperFlow i instrukcja jak definiować workflow)
- <https://github.com/dice-cyfronet/hyperflow-client>
- <https://github.com/dice-cyfronet/hyperflow-amqp-executor>
- <http://www.rabbitmq.com/>