Usługa Powszechnej Archiwizacji - PLATON

- Krótki opis usługi
- Aktywowanie usługi
  - Ograniczenia w korzystaniu
- Pierwsze kroki
  - Programy klienckie
  - Schemat integracji usługi z PL-Grid
  - Punkty dostępowe PLATON
  - Katalog domowy użytkownika
  - Dostęp do usługi przez program sftp
  - Dostęp do usługi przez program sshfs
  - Dostęp do usługi przez program globus-url-copy

Krótki opis usługi

Usługa Powszechnej Archiwizacji – PLATON (nazywana dalej w skrócie usługą PLATON lub usługą powszechnej archiwizacji), jest adresowana do polskiego środowiska naukowego i służy do składowania kopi zapasowych i archiwalnych dużych wolumenów danych. W przypadku użytkowników portalu PL-Grid, wgrane dane są utrzymywane w dwóch kopiach, w rozproszonych geograficznie ośrodkach KDM. Użytkownik korzysta z usługi przez protokół sftp lub GridFTP. Dostęp do usługi możliwy jest z poziomu jednego z węzłów dostępowych PL-Grid (tj. z UI). W PLATON-ie serwery pozwalające na dostęp do usługi nazywane są punktami dostępowymi. Punkty dostępowe PLATON są rozmieszczone w tych samych ośrodkach co węzły dostępowe PL-Grid. Każdy punkt dostępowy PLATON eksponuje dokładnie tą samą przestrzeń danych. Pomijając kwestie wydajnościowe, bez znaczenia jest, za pośrednictwem którego z punktów dostępowych PLATON, użytkownik korzysta z usługi. Punkty dostępowe PLATON składują dane użytkowników w punktach składowania. Lokalizacja punktów składowania danych również pokrywa się z lokalizacją węzłów dostępowych PL-Grid. Docelowe punkty składowania danych są wybierane przez inteligentne algorytmy realizowane przez punkty dostępowe PLATON. Przy wyborze punktu składowania danych uwzględnia się takie parametry jak lokalizacja użytego punktu dostępowego PLATON oraz chwilowa dostępność punktów składowania w potencjalnych docelowych lokalizacjach. Dane użytkowników portalu PL-Grid są replikowane w trybie asynchronicznym, tj. gdy użytkownik wgrywa plik, najpierw trafia on do możliwie najbliższego punktu składowania danych, a następnie, gdy wgrywanie się zakończy, to w tzw. tle tworzona jest jego kopia w punkcie składowania położonym w innej lokalizacji. Głównym celem rozproszenia geograficznego danych jest zwiększenie bezpieczeństwa i dostępności danych.

Istotną cechą systemu jest fizyczny sposób składowania danych. Z uwagi na koszt jednostkowy przechowywania dużych wolumenów danych, w tym konkretnym wdrożeniu usługi PLATON, dane użytkowników, ostatecznie trafiają na bibliotekę taśmową. Oznacza to, że późniejsza próba dostępu do danych może się wiązać nawet z kilku minutowym opóźnieniem związanym z pracą tzw. robota ładującego taśmy oraz z oczekiwaniem na dostęp do wolnego (tj. nie zajętego przez inne taśmy) czytnika taśm. Jest to właściwość, która dość jasno determinuje przeznaczenie i scenariusze wykorzystania usługi (tworzenie archiwów i kopii zapasowych, składowanie dużych plików, rzadki odczyt danych).

Aktywowanie usługi

Usługa jest dostępna poza systemem przydziału grantów, dlatego jest dostępna dla każdego użytkownika portalu i nie wymaga specyfikacji w grantach.

Aby aktywować usługę, postępuj zgodnie z instrukcjami zawartymi w Katalogu Aplikacji i Usług po uprzednim wyszukaniu usługi. Wygeneruj też certyfikat Simple CA lub zarejestruj certyfikat Polish Grid CA (patrz: Certyfikaty).

Gdy usługa zostanie aktywowana, to dostęp do niej będzie możliwy z poziomu węzłów dostępowych PLGrid wymienionych poniżej w sekcji "Punkty dostępowe PLATON".

Ograniczenia w korzystaniu

Każdy użytkownik może przechowywać w usłudze PLATON do 100GB danych.

Z usługi PLATON korzysta się za pośrednictwem typowych programów klienckich obsługujących protokół sftp lub GridFTP. Ponieważ do uwierzytelniania usługa używa certyfikatów (w przypadku protokołu GridFTP) i kluczy (w przypadku protokołu sftp), to przed aktywowaniem usługi, najpierw z poziomu portalu PL-Grid, należy wygenerować a następnie zarejestrować w portalu certyfikat SimpleCA. Możliwe jest też zarejestrowanie certyfikatu Polish Grid CA (jeśli akurat ktoś taki posiada). Jest to istotne, ponieważ dopiero po rejestracji certyfikat i powiązane z nim klucze automatycznie będą widoczne na węzłach dostępowych PL-Grid.

Dostęp do usługi (czyli do punktów dostępowych PLATON) możliwy jest tylko z poziomu serwerów dostępowych PL-Grid. Zakres wykorzystania usługi ograniczony jest do składowania kopii zapasowych i archiwów. Mimo, że dzięki programowi sshfs, przestrzeń dostępną w ramach usługi można podmontować w lokalnym katalogu, to nie należy korzystać z dostarczonej w ten sposób przestrzeni jak z typowego dysku sieciowego (nie wgrywamy dużej ilości małych plików, nie edytujemy plików otwartych bezpośrednio z tej przestrzeni, nie pozwalamy aplikacjom obliczeniowym umieszczać bezpośrednio w tej przestrzeni końcowych lub pośrednich plików wynikowych).

Teoretycznie, protokoły dostępowe sftp i gridftp pozwalają na wykonanie na zdalnych plikach operacji zmiany grupy będącej właścicielem danego pliku (operacja chmod). Standardowo, w systemie Linux, jeśli użytkownik jest właścicielem pliku, to może zmieniać atrybut pliku określający grupę będącą właścicielem danego pliku. Jest to jednak możliwe tylko w zakresie grup do których użytkownik sam należy. Infrastruktura PL-Grid zapewnia, że użytkownicy portalu PL-Grid mają automatycznie zakładane konta na węzłach dostępowych. Na każdy węzeł można się zalogować przy pomocy tej samej nazwy użytkownika i hasła, których używa się w portalu. Trzeba mieć na uwadze, że w środowisku PL-Grid na każdym z węzłów dostępowych UID przydzielony użytkownikowi jest inny, brak jest możliwości wykonania operacji *chmod*.

Z punktami dostępowymi usługi PLATON (tzw. AN) zawsze łączymy się z poziomu jednego z węzłów dostępowych PL-Grid (tzw. UI). Jeśli połączenia będą nawiązywane z najbliższym punktem dostępowym, to poprawi to wydajność dostępu do danych. Tabela zestawiająca lokalizacje punktów AN i węzłów UI znajduje się w dalszej części dokumentu.

Pierwsze kroki

Programy klienckie

Dostęp do usługi może się odbywać tylko z poziomu serwerów dostępowych PL-Grid, za pośrednictwem typowych programów klienckich obsługujących protokoły sftp i gridftp (program sftp, sshfs. globus-url-copy, uberftp). Programem klienckim należy się połączyć z jednym z pięciu punktów dostępowych PLATON (lista adresów poniżej). Do uwierzytelniania użytkowników, usługa używa certyfikatów lub kluczy. Dostępność potrzebnych do uwierzytelnienia certyfikatów lub kluczy zapewniana jest przez infrastrukturę PL-Grid (dokładnie przez wirtualny system plików keyfs).

Schemat integracji usługi z PL-Grid

Na wstępie należy zaznaczyć różnice pomiędzy terminami: węzeł dostępowy, punkt dostępowy, punkt składowania danych i lokalizacja.

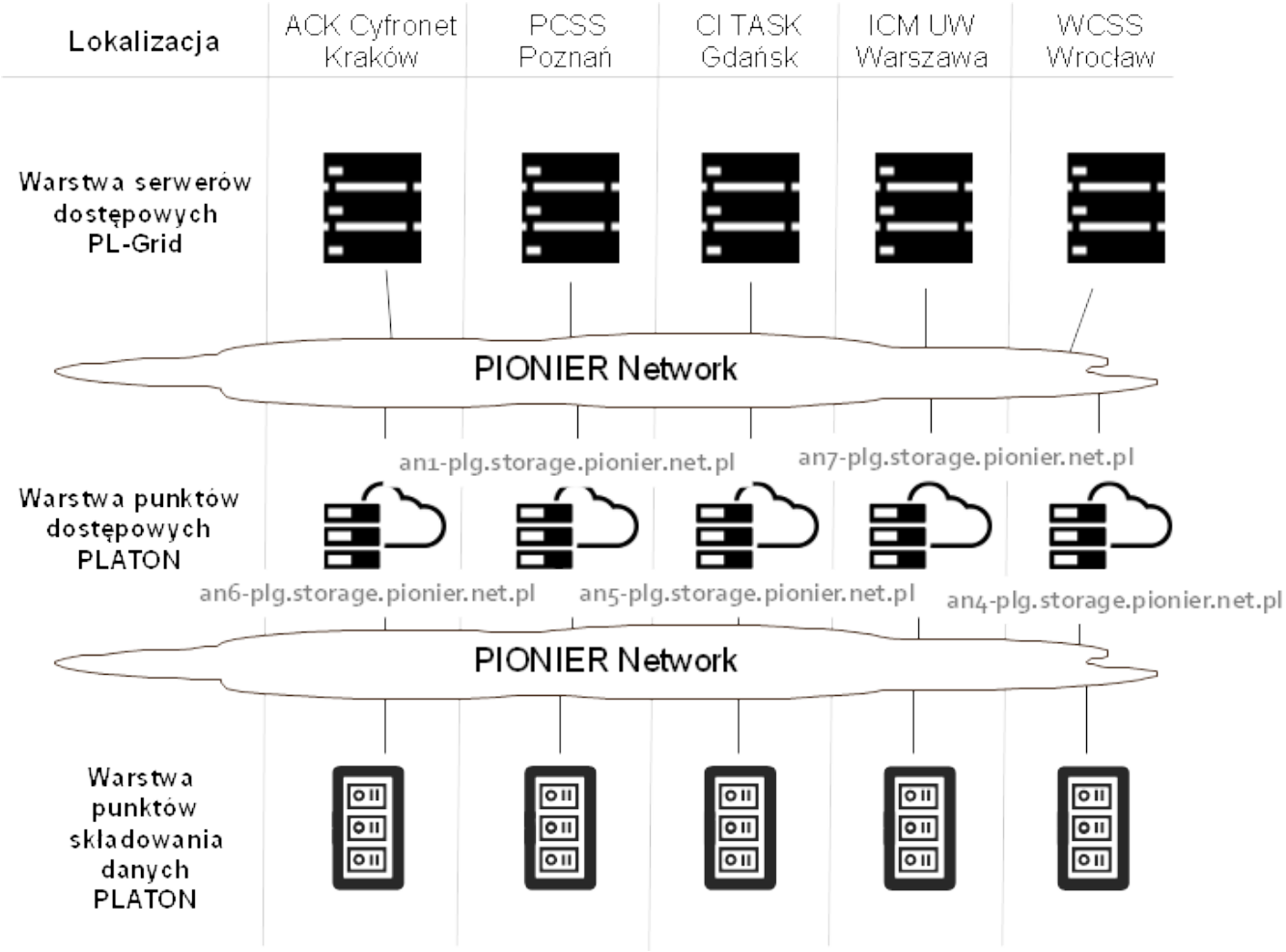
Przez **węzeł dostępowy** lub **serwer dostępowy** rozumiemy jeden z serwerów infrastruktury PL-Grid, wymieniony w punkcie "Punkty dostępowe PLATON" w tabeli zestawiającej węzły PL-Grid z najbliższym punktami dostępowymi PLATON.

Punkt dostępowy to z kolei jeden z serwerów infrastruktury Usługi Powszechnej Archiwizacji - PLATON.

Punkt składowania danych to wykorzystywany przez usługę PLATON system informatyczny przeznaczony do bezpiecznego, długoterminowego przechowywania dużych wolumenów danych.Precyzyjne określenie położenia i technologii nie jest istotne, bo z założenia jest przezroczyste dla użytkowników końcowych.

Najbardziej ogólnym pojęciem jest **lokalizacja**, która oznacza jedynie przybliżone geograficzne odniesienie do miejsca w którym znajduje się ośrodek, w którym są umiejscowione serwery węzłów lub punktów dostępowych oraz punkty składowania danych.

Poniższy rysunek przedstawia relację między węzłami dostępowymi, punktami dostępowymi oraz punktami składowania danych.



Na rysunku widać, że zarówno punkty dostępowe PLATON jak i docelowe punkty składowania danych znajdują się w tych samych ośrodkach (lokalizacjach) co węzły dostępowe PL-Grid. Podczas korzystania z usługi, wybór punktów składowania jest dla użytkownika przezroczysty. Najczęściej, jeśli będzie to możliwe, pliki będą składowane w lokalizacji najbliższej w stosunku do użytego punktu dostępowego PLATON. Druga kopia pliku jest tworzona w lokalizacji wybieranej wg wewnętrznych algorytmów systemu.

Punkty dostępowe PLATON

Z usługi PLATON można korzystać za pośrednictwem dowolnego punktu dostępowego PLATON, jednak dla zwiększenia wydajności systemu, zaleca się by był to punkt dostępowy najbliższy wykorzystywanemu serwerowi dostępowemu PL-Grid. Poniższa tabela przedstawia punkty dostępowe PLATON w zestawieniu z najbliższymi serwerami dostępowymi PL-Grid.

Lokalizacja	Nazwa węzła dostępowego PL-Grid	Nazwa punktu dostępowego PLATON
ACK Cyfronet, Kraków	ui.cyfronet.pl	an6-plg.storage.pionier.net.pl
PCSS, Poznań	ui.reef.man.poznan.pl	an1-plg.storage.pionier.net.pl
CI TASK, Gdańsk	ui.grid.task.gda.pl	an5-plg.storage.pionier.net.pl
ICM UW, Warszawa	ui.plgrid.icm.edu.pl login.icm.edu.pl	an7-plg.storage.pionier.net.pl
WCSS, Wrocław	ui.plgrid.wcss.wroc.pl ui.wcss.pl	an4-plg.storage.pionier.net.pl

Tabela: Zestawienie punktów dostępowych PLATON i węzłów dostępowych PL-Grid dostępnych w poszczególnych lokalizacjach.

Katalog domowy użytkownika

W usłudze PLATON, dostępnej dla użytkowników portalu PL-Grid, każdy użytkownik ma swój katalog domowy. Prawo dostępu do tego katalogu ma tylko jego właściciel (tj. użytkownik dla którego katalog został utworzony).

Katalogi domowe użytkowników w usłudze PLATON znajdują się w katalogu **/home/kmd/plgrid-portal/users**.

Programy takie jak sfpt czy sshfs, domyślnie realizują dostęp do katalogu domowego, więc w przypadku tych programów nie ma potrzeby jawnego podawania ścieżki do katalogu domowego.

Korzystając z protokołu GridFTP i programu klienckiego globus-url-copy, wymagane będzie jawne podanie pełnej ścieżki do katalogu domowego (przykład w dalszej części opisu).

Dostęp do usługi przez program sftp

Procedura dostępu do usługi poprzez program sftp:

1. Logujemy się na jeden z serwerów dostępowych PL-Grid.
2. Używając jednego z punktów dostępowych PLATON (patrz tabela Punkty dostępowe PLATON) oraz klucza dostarczanego przez keyfs i dostępnego pod ścieżką **/mnt/keyfs/users/<username>/.globus/userkey.pem**, nawiązujemy połączenie z usługą:

```
sftp -o IdentityFile=/mnt/keyfs/users/<username>/.globus/userkey.pem <platon_access_node>
```

gdzie,

-o IdentityFile to parametr wskazujący klucz prywatny użytkownika,

<username> to nazwa użytkownika PL-Grid,

<platon_access_node> to nazwa punktu dostępowego PLATON, np. *an6-plg.storage.pionier.net.pl*

UWAGA: Klucz prywatny użytkownika jest chroniony hasłem ustawionym z poziomu portalu PL-Grid. Program sftp, podczas uwierzytelniania w usłudze, poprosi nas o podanie tego hasła.

3. Po poprawnym uwierzytelnieniu, program sftp dostarcza standardowego interfejsu w postaci tekstowej konsoli sftp. Domyślnie, bezpośrednio po uzyskaniu dostępu, zdalnym katalogiem roboczym jest katalog domowy użytkownika. Do wyświetlenia listy dostępnych komend służy polecenie **help**.

```
sftp> help
Available commands:
bye Quit sftp
cd path Change remote directory to 'path'
chgrp grp path Change group of file 'path' to 'grp'
chmod mode path Change permissions of file 'path' to 'mode'
chown own path Change owner of file 'path' to 'own'
df [-hi] [path] Display statistics for current directory or
filesystem containing 'path'
exit Quit sftp
get [-P] remote-path [local-path] Download file
help Display this help text
lcd path Change local directory to 'path'
lls [ls-options [path]] Display local directory listing
lmkdir path Create local directory
ln oldpath newpath Symlink remote file
lpwd Print local working directory
ls [-1aflnrSt] [path] Display remote directory listing
lumask umask Set local umask to 'umask'
mkdir path Create remote directory
progress Toggle display of progress meter
put [-P] local-path [remote-path] Upload file
pwd Display remote working directory
quit Quit sftp
rename oldpath newpath Rename remote file
rm path Delete remote file
rmdir path Remove remote directory
symlink oldpath newpath Symlink remote file
version Show SFTP version
!command Execute 'command' in local shell
! Escape to local shell
? Synonym for help
```

4. Przykład wgrania do usługi PLATON pliku /etc/profile:

```
sftp> put /etc/profile
Uploading /etc/profile to /home/kmd/plgrid-portal/users/plggracjan/profile
/etc/profile 100% 1796 1.8KB/s 00:00
```

5. Przykład wyświetlenia listy plików wgranych do usługi:

```
sftp> ls
profile
```

6. Przykład pobrania z usługi pliku profile. Plik zostanie zapisany na UI w bieżącym katalogu roboczym:

```
sftp> get profile
Fetching /home/kmd/plgrid-portal/users/plggracjan/profile to profile
/home/kmd/plgrid-portal/users/plggracjan/prof 100% 1796 1.8KB/s 00:00
```

7. Przykład usunięcie pliku profile z usługi:

```
sftp> rm profile
Removing /home/kmd/plgrid-portal/users/plggracjan/profile
```

8. Zakończenie pracy z sfpt:

```
sftp> exit
```

Dostęp do usługi przez program sshfs

Program sshfs pozwala na montowanie w lokalnym katalogu zasobu dostępnego poprzez protokół sftp. Przykładowa procedura użycia tego programu wygląda następująco:

1. Logujemy się na jeden z serwerów dostępowych PL-Grid.
2. W katalogu domowym (na serwerze dostępowym PL-Grid) zakładamy katalog pomocniczy, do którego następnie montujemy przestrzeń udostępnianą przez usługę PLATON.

```
mkdir platon-backup
```

3. Używając jednego z punktów dostępowych usługi PLATON (patrz tabela Punkty dostępowe PLATON) oraz klucza dostarczonego przez keyfs i dostępnego pod ścieżką **/mnt/keyfs/users/<username>/.globus/userkey.pem**, montujemy przestrzeń wystawianą przez usługę w lokalnym katalogu (w tym przypadku w katalogu platon-backup):

```
sshfs -o idmap=user,gid=$(id -g) -o IdentityFile=/mnt/keyfs/users/<username>/.globus/userkey.pem
<platon_access_node>:. platon-backup
```

gdzie,

-o idmap=user to parametr włączający mapowanie UID-ów właściciela plików po stronie serwera ssh na UID-y lokalnego użytkownika, jest to parametr pomagający trochę ukryć niedogodności wynikające z faktu że w PL-Grid ten sam użytkownik ma różne UID-y na różnych węzłach dostępowych

-o IdentityFile to parametr wskazujący ścieżkę do klucza prywatnego użytkownika,

<username> to nazwa użytkownika PL-Grid,

<platon_access_node> to nazwa punktu dostępowego usługi PLATON

UWAGA: Klucz prywatny użytkownika jest chroniony hasłem ustawionym z poziomu portalu PL-Grid. Program sftp, podczas uwierzytelniania w usłudze, poprosi nas o podanie tego hasła.

4. Po poprawnym zakończeniu powyższego polecenia, przestrzeń udostępniana przez usługę PLATON powinna być dostępna na serwerze dostępowym PL-Grid w katalogu domowym użytkownika w podkatalogu platon-backup.
5. Wgrywanie, pobieranie i usuwanie plików z usługi PLATON, sprowadza się do standardowych operacji na systemie plików.
6. Ewentualnego odmontowania przestrzeni dokonujemy poleceniem:

```
fusermount -u platon-backup
```

gdzie,

platon-backup to ścieżka do katalogu w którym podmontowaliśmy wcześniej przestrzeń wystawioną przez usługę PLATON, w tym przypadku jest to ścieżka względna, bezpośrednio wskazująca na katalog w bieżącym katalogu roboczym

UWAGA: Z uwagi na charakterystykę wirtualnych systemów plików wykorzystywanych do świadczenia usługi PLATON, polecenie **df** nie będzie pokazywać poprawnego wykorzystania zasobów w punktach podmontowanych z tej usługi.

Dostęp do usługi przez program globus-url-copy

Opis wykorzystania protokołu GridFTP można znaleźć na stronie Transfer danych - GridFTP.

Przykładowy scenariusz użycia, z wykorzystaniem klienta globus-url-copy, może wyglądać tak:

1. Logujemy się na jeden z serwerów dostępowych PL-Grid.
2. Korzystając z certyfikatu i klucza wystawionego przez keyfs, generujemy tzw. certyfikat proxy, używany przez GridFTP do uwirzytelniania użytkowników:

```
grid-proxy-init
```

UWAGA: Powyższe polecenie generuje certyfikat proxy bazując na pliku z certyfikatem i kluczem prywatnym użytkownika. Lokalizacja plików z certyfikatem i kluczem, jest ustalana na podstawie wartości zmiennych środowiskowych X509_USER_CERT i X509_USER_KEY, które na węźle dostępowym PL-Grid są automatycznie tak ustawiane żeby wskazywały odpowiednie pliki w wirtualnym systemie plików keyfs.

3. Domyślnie certyfikat proxy ma ważność 12 godzin. Po wygaśnięciu certyfikatu należy wygenerować nowy. Do sprawdzenia czasu ważności certyfikatu proxy służy polecenie:

```
grid-proxy-info
```

4. Do kopiowania plików służy komenda globus-url-copy. Przykładowe przekopiowanie pliku z serwera dostępowego PL-Grid do usługi PLATON może wyglądać tak:

```
globus-url-copy file:///etc/issue gsiftp://an1-plg.storage.pionier.net.pl/home/kmd/plgrid-portal
/users/$(whoami)/test
```

gdzie,

file:///etc/issue to URL do lokalnego pliu /etc/issue,

an1-plg.storage.pionier.net.pl to adres jednego z punktów dostępowych PLATON,

$(whoami) to wywołanie polecenia zwracającego nazwę bieżącego użytkownika na UI, pozwala to w uogólniony sposób odwołać się do katalogu domowego utrzymywanego dla danego użytkownika po stronie usługi PLATON.

5. Analogicznie, przegranie pliku z usługi PLATON na węzeł UI, wygląda tak:

```
globus-url-copy gsiftp://an1-plg.storage.pionier.net.pl/home/kmd/plgrid-portal/users/$(whoami)/test
file:///tmp/test-kopia
```

Jest to skopiowanie pliku test z usługi PLATON do lokalnego katalogu tmp. Nowa kopia pliku ma nazwę "test-kopia".

6. Przykład wylistowania plików z katalogu domowego w usłudze PLATON:

```
globus-url-copy -list gsiftp://an1-plg.storage.pionier.net.pl/home/kmd/plgrid-portal/user/$(whoami)
```