
Gestion des données sur la grille

Version 3.0

Jérôme Pansanel

août 23, 2023

Table des matières

1	Introduction	1
2	Les systèmes de stockage HTC	3
3	Gestion des données avec le service HTC	5
3.1	Généralités	5
3.2	Installation du client	5
3.3	Lecture du contenu d'un répertoire	5
3.4	Création d'un répertoire	6
3.5	Copie d'un fichier	6
3.6	Suppression d'un fichier	6
3.7	Suppression d'un répertoire	6
4	Documentation complémentaire	7

CHAPITRE 1

Introduction

La [plateforme SCIGNE](#) propose un service de calcul à haut débit (*high-throughput computing*, HTC) reposant sur des technologies développées pour les grilles de calcul. Les échanges de données entre les différents centres de calcul composants ces grilles sont réalisés à l'aide de protocoles spécifiques (HTTP et XROOTD).

Ce document détaille l'utilisation de l'outil GFAL2 afin d'interagir avec ces protocoles pour gérer vos données depuis votre station de travail, la plateforme SCIGNE ou l'un des sites partenaires.

Les systèmes de stockage HTC

Plusieurs systèmes de stockage (DPM, dCache, Storm, etc) sont utilisés par les sites des grilles de calcul. Dans le but d'assurer l'interopérabilité des systèmes entre eux, ils communiquent avec des protocoles identiques.

La plateforme SCIGNE utilise le système dCache pour mettre à disposition le stockage HTC. Bien que plusieurs protocoles soient disponibles (HTTP, XROOTD), cette documentation se concentre sur l'utilisation de HTTP.

Le système de stockage dCache est composé d'un serveur de tête et d'un ensemble de serveurs de fichiers. Le serveur de tête centralise les requêtes d'accès aux fichiers. Les serveurs de fichiers stockent physiquement les fichiers. L'accès aux fichiers s'effectue en interrogeant le serveur de tête à l'aide des commandes détaillées dans ce document.

L'accès aux données est contrôlé par certificat. Il est donc nécessaire de posséder un *proxy* valide lors de l'utilisation de ce type de stockage.

Gestion des données avec le service HTC

Cette section détaille la gestion des données stockées sur un serveur dCache.

3.1 Généralités

L'URL suivante est utilisée pour l'ensemble de cette section. Elle sera abrégée en `${BASE_URL}` :

```
$ export BASE_URL="davs://sbgdcache.in2p3.fr/vo.grand-est.fr"
```

3.2 Installation du client

Actuellement, les paquets ne sont disponibles que pour la distribution RedHat et ses dérivés. Les paquets sont disponibles dans le dépôt EPEL. Une fois ce dépôt disponible sur votre système, les outils `gfal2` et peuvent être rapidement installés avec :

```
$ yum install gfal2-util gfal2-all
```

3.3 Lecture du contenu d'un répertoire

La commande `gfal-ls` permet de lister le contenu d'un répertoire. Le résultat de cette commande est similaire à celui de la commande `ls`. L'exemple ci-dessous illustre le fonctionnement de cette commande :

```
$ gfal-ls ${BASE_URL}
EOST
LIMH
LIVE
UMR7200
ICS
```

```
$ gfal-ls -l ${BASE_URL}
drwxrwxrwx 0 0 0 0 Aug 22 07:42 EOST
drwxrwxrwx 0 0 0 0 Aug 22 07:42 LIMH
drwxrwxrwx 0 0 0 0 Aug 22 07:42 LIVE
drwxrwxrwx 0 0 0 0 Aug 22 07:42 UMR7200
drwxrwxrwx 0 0 0 0 Aug 22 07:43 ICS
```

Le contenu de ce répertoire évoluant dans le temps, le résultat affiché par les commandes précédentes diffère des exemples ci-dessus.

3.4 Création d'un répertoire

La commande `gfal-mkdir` permet de créer un répertoire :

```
$ gfal-mkdir ${BASE_URL}/lab/user/data
```

Notez que la valeur `/lab/user/data` est donnée à titre d'exemple et doit être adaptée à votre cas d'usage.

3.5 Copie d'un fichier

La commande `gfal-copy` permet de copier des données depuis ou vers le serveur dCache :

```
$ gfal-copy file://${PWD}/my_dataset.tar.gz ${BASE_URL}/lab/user/data/my_dataset.tar.
↪gz
Copying file://${PWD}/my_dataset.tar.gz [DONE] after 2s
```

3.6 Suppression d'un fichier

La commande `gfal-rm` permet de supprimer un fichier sur le serveur dCache :

```
$ gfal-rm ${BASE_URL}/lab/user/data/my_dataset.tar.gz
davs://sbgdcache.in2p3.fr/vo.grand-est.fr/lab/user/data DELETED
```

3.7 Suppression d'un répertoire

La commande `gfal-rm -r` permet de supprimer un répertoire (et son contenu) :

```
$ gfal-rm -r ${BASE_URL}/lab/user/data
```

CHAPITRE 4

Documentation complémentaire

Ces liens proposent des références complémentaires pour approfondir la gestion des données sur la plateforme SCIGNE :

- [documentation sur les commandes gfa12](#);
- [documentation sur dCache](#);