

スパコン計算データリポジトリの立ち上げ

鈴木和磨[#]

自然科学研究機構 分子科学研究所

概要

近年、オープンサイエンスや分野横断的な研究を推進する流れを受けて、研究者が出版社や他の研究室などの外部の人に対してデータを共有・公開する機会が増加している。しかし、一般的なファイル共有サービスは、大容量の計算データのやり取りに用いるには扱えるデータ容量が少ない上、データが長期間保存される保証がない。この課題に対して、当センターではスーパーコンピューター利用者が無料で利用可能な独自の計算データリポジトリサービスを立ち上げることとなった。本発表では、本サービスの要件を満たすソフトウェアの調査・選定や運用方針の決定など、サービス開始に向けて現在までにどのような取り組みを行ってきたか発表する。

1. 立ち上げの背景

近年、オープンサイエンスや分野横断的な研究を推進する流れを受けて、研究者が出版社や他の研究室などの外部の人に対してデータを共有・公開する機会が増えている。それに伴い、データを保存し、インターネット上で共有および公開するためのプラットフォームであるデータリポジトリのニーズが増加している。データリポジトリの代表例が Zenodo[1]である。Zenodo は OpenAIRE と欧州原子核研究機構 (CERN) によって作られた汎用リポジトリであり、どんな研究分野の研究者もファイルをアップロードできる。また、オープンアクセスリポジトリ推進協会 (JPCOAR) と国立情報学研究所 (NII) が共同運営する JAIRO Cloud[2]は、大学や研究機関が運営する機関リポジトリの構築環境を提供するサービスである。機関リポジトリは、博士論文などの公開が中心ではあるが研究データの公開にも対応している場合がある。

こうした外部のデータリポジトリサービスを利用する上で気になるのはアップロードできるファイルサイズの上限である。上に挙げたサービスの場合、Zenodo には 50 GB、JAIRO Cloud には 100 MB までのサイズのファイルをアップロードすることができるが、100 GB を超える大規模なデータを扱うこともあるスパコンユーザーにとってこの上限は小さい。この課題に対して、計算科学研究センターではスパコンユーザー向けの計算データリポジトリを独自に立ち上げることとした。

2. スパコン計算データリポジトリの要件

独自で計算データリポジトリを立ち上げるうえで挙げられた要件を箇条書きで示す。

- 利用者はセンターのスパコンを用いて成果を上げているグループを想定している。つまり誰にでも利用させるわけではない。
- 主に論文投稿時に共有・公開を要求されるデータの置き場として使用したいのでデータへのリンク生成ができる必要がある。
- 研究グループ間でデータのやり取りにも使用し

たい。

- ファイル容量としては1グループあたり1TBほしい。
- ファイルのアップロードがドラッグ&ドロップでできるなど操作はできるだけ簡単であることが望ましい。
- データの公開範囲は利用者が設定できる。
- 検索機能は必要ない。

3. 要件を満たすソフトの調査:Nextcloud

簡単な操作でファイルのアップロードや共有ができるソフトを探した結果、オープンソースのオンラインストレージ構築ソフト Nextcloud にたどり着いた。Nextcloud はドラッグアンドドロップによるファイルアップロードの他、WebDAV を用いたコマンドライン操作によるファイル移動が可能である。スパコンユーザーにとっては操作がコマンドラインで完結するのは嬉しい要素である。また、Nextcloud ではフォルダ公開の際はフォルダへアクセスするための共有リンクを生成する。リンクのパーミッションや公開期限はカスタマイズ可能である。

Nextcloud を仮想サーバー上にインストールし、実際に機能と使用感を検証した結果、Google Drive のような簡単な操作感でファイルアップロードできることがわかった。しかも共有リンクの生成に関しては Google Drive 以上にわかりやすく簡単であると感じた。要件も満たしているため、スパコン計算データリポジトリとして Nextcloud を採用することを決めた。

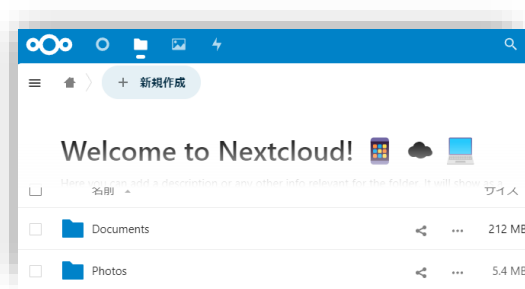


図 1. Nextcloud 画面

4. Nextcloud 運用方針の決定

要件に基づき Nextcloud の運用方針を決定した。まず、ユーザーの多くがスパコン上からデータをアップロードすることを想定し、Nextcloud はスパコンのフロントエンドサーバー上にインストールすることにした。これによりスパコン上から巨大なデータを比較的短い時間でアップロードすることが可能となる。アップロードファイル本体はスパコンにマウントされた分散ファイルシステム Lustre に置くこととした。Lustre は 14.8 PB のディスク容量がある。

アカウントはグループにつき1つとし、作成にはセンター長による審査・承認が必要とする。アカウントは研究室の主催者 (PI) に紐づけるが、申請自体は PI 以外も可能とする。ただし、PI も当センターのスパコン利用者でなければならない。要件通り、ディスク容量は1グループにつき1TBに設定した。ただし、サービス利用状況によっては変更する可能性がある。Lustre 上のデータについてはバックアップを取らない方針であるため、Nextcloud 上のデータのバックアップについてはユーザーの責任とする。

メンテナンスは、スパコンのメンテナンス日と同日とし、その際にソフトウェアのアップデートなどを行うものとする。Nextcloud の機能であるメンテナンスモードに切り替えることでデータのアップロードなどをロックする。

セキュリティ対策として、ユーザーには二段階認証を推奨する。認証方法としてメールか Time-based One-time Password (TOTP) を使う方法が選択できる。これらは Nextcloud の拡張機能として提供されている。

5. 本番環境へのインストールと設定

Nextcloud を本番環境であるスパコンのフロントエンドサーバー上にインストールした。インストールには、ウェブサーバー、php バージョン 8.1 以降およびデータベースが必要である。ウェブサーバーには Apache、php には remi リポジトリから取得したバージョン 8.2、データベースには MariaDB を採用した。大容量のファイルをアップロードできるように、`/etc/php.ini` を編集してアップロード可能なファイルサイズの上限をデフォルトの 512 MB から増やした。フロントエンドサーバーに WebDAV クライアントである cadaver をインストールし、コマンドライン操作によるファイルアップロードができるようにした。

6. Nextcloud 利用開始までの流れ

現在は、サービス公開に向けてユーザーが利用申請してからアカウントを作成するまでの流れを整備している。登録フローを図2に示す。

申請開始～DB登録

このステップでは申請者が情報入力するフォームが必要である。図3に作成した申請フォームを示す。この申請フォームは計算科学研究センターのホームページである ccportal[3]上のスパコン利用者限定ページに掲載する予定である。ccportal は CMS である



図2. Nextcloud 利用開始までの流れ

Drupal によって作成・管理されているため、申請フォームは Drupal のモジュールとして作成した。入力項目はメールアドレス、リポジトリのアカウント名 (6文字以上 30文字以下)、200字以内の利用目的、(申請者が PI でなければ) PI のスパコンユーザーID である。Nextcloud のアカウント名は英数字、ダッシュ、アンダースコア、ピリオド、スペース、@のみが使用可能であるため、それ以外が入力されていると申請できないようになっている。アカウント名が既に DB に登録されているユーザーのものと同様に被っていてもエラーとなる。フォームは英語に対応しており、英語版では利用目的は 100 単語以下で書いてもらうようになっている。利用目的については自作の JavaScript を用いて、あと何文字 or 何単語入力できるか表示されるようになっている。申請ボタンを押して、入力内容に問題がなければ審査前ユーザーとして DB に登録され、申請者と管理者にメールが送られる。

DB登録～審査待ち

審査前ユーザーの一覧は自作の管理インターフェース (図4) で確認できる。ID列のリンクをクリックすると審査リクエスト確認画面 (図5) に遷移する。画面には申請者が入力した内容と、スパコン利用者の DB から取得した名前、所属、役割が表示される。この情報を添えて担当者に審査依頼を送り、Review request ボタンを押す。ボタンを押すとユーザーステータスが審査中に変化する。

審査待ち～アカウント作成・通知

審査中ユーザーは図4のインターフェースの別タブから確認できる。UI は審査前と共通で ID をクリックすると登録確認画面に遷移する。審査の結果、アカウント作成が承認され、確認画面の Registration ボタンを押すとアカウント作成スクリプトが動く。拒否の場合は、Reject ボタンを押してその旨を申請者にメールで伝える。アカウント作成スクリプトは Nextcloud の REST API を利用する。アカウント作成が正常に完了した場合は、初期パスワードと共に Nextcloud が利用可能となった旨を申請者に伝える。

7. 今後の課題

現在、当センターの先生方にテスト利用をいただいている。大きな問題なく利用できているようだが、利用者が増えたときに備えてパフォーマンスをチューニングする方法を学ぶ必要がある。

参考文献

- [1] <https://zenodo.org/>
- [2] <https://jpcoar.repo.nii.ac.jp/page/42>
- [3] <https://ccportal.ims.ac.jp/>

Application form of the RCCS data repository

PIでない方が申請する場合、以下のチェックボックスにチェックを入れ、PIの方のユーザーIDを入力してください。

私はPIではありません。

PIのユーザーID *

メールアドレス *

リポジトリのアカウント名 (6文字以上30文字以下) *

英数字(a-z, A-Z, 0-9)、ダッシュ(-)、アンダースコア(_)、ピリオド(.)、スペース()、@が使用可能です。

利用目的 (200字以内) *

残り196文字

▶ 利用規約

利用規約に同意します。 *

申請

図 3. Nextcloud 申請フォーム

Users before review

Users before review | Users under review

ID	PI's name	Workplace and Department	Position	Repository ID
1	鈴木 和磨	分子科学研究所 技術推進部計算情報ユニット	技術職員	ksuzuki_ims

図 4. ユーザー管理インターフェース

Confirm Request

Application date	2024-02-01
Repository user ID	ksuzuki_ims
Name	鈴木 和磨
E-mail	ksuzuki@ims.ac.jp
Affiliation	分子科学研究所 技術推進部計算情報ユニット
Position	技術職員
Purpose of use	test

Review request

図 5. 審査リクエスト確認画面