

金沢大学事務用アーカイブ領域の導入

○松本 好美、浜 貴幸
金沢大学総合技術部情報部門

1. はじめに

金沢大学では、事務職員がデータファイルを共有するためのストレージシステムを事務用情報システムとして導入しており、共有フォルダとして利用者に提供している。共有フォルダのデータ保存量は年々増加しており、部署毎に設定されたクォータの警告通知が出ると利用者には不要なデータの削除依頼を行い、データの削除ができない場合はクォータを拡張して対応していた。しかし、異動が多い事務職員は過去のデータを整理することが難しく、データの整理が進まないままクォータの拡張を求められることが多い。

そのため、2023年の事務用情報システム更改^[1]時には共有フォルダのストレージ容量の拡張を行うと共にテープストレージの導入を行い、使用頻度は低いけど削除できないデータを保管する「アーカイブ領域」として運用を開始した。ここでは、このアーカイブ領域の導入と運用について報告する。

2. システム構成

事務用情報システムは主に4つのサブシステムから構成されている(図1)。

リモートデスクトップ接続ブローカーや Active Directory サーバなどシステム稼働に必要な仮想サーバ群と NVMe SSD で構成されたストレージ装置からなる基幹サーバシステム、利用者がリモートデスクトップ接続する RDSH (Remote Desktop Session Host) 42 式からなる RDSH クラスタシステム、利用者がリモートデスクトップ接続するクライアント端末として使用するノート PC など含むターミナルシステム、そしてバックアップストレージ装置及びテープドライブ装置から構成するニアラインストレージシステムである。

今回導入したアーカイブ領域は基幹サーバシステムのストレージ装置とニアラインストレージサーバ及びテープドライブ装置の連携により実現している。

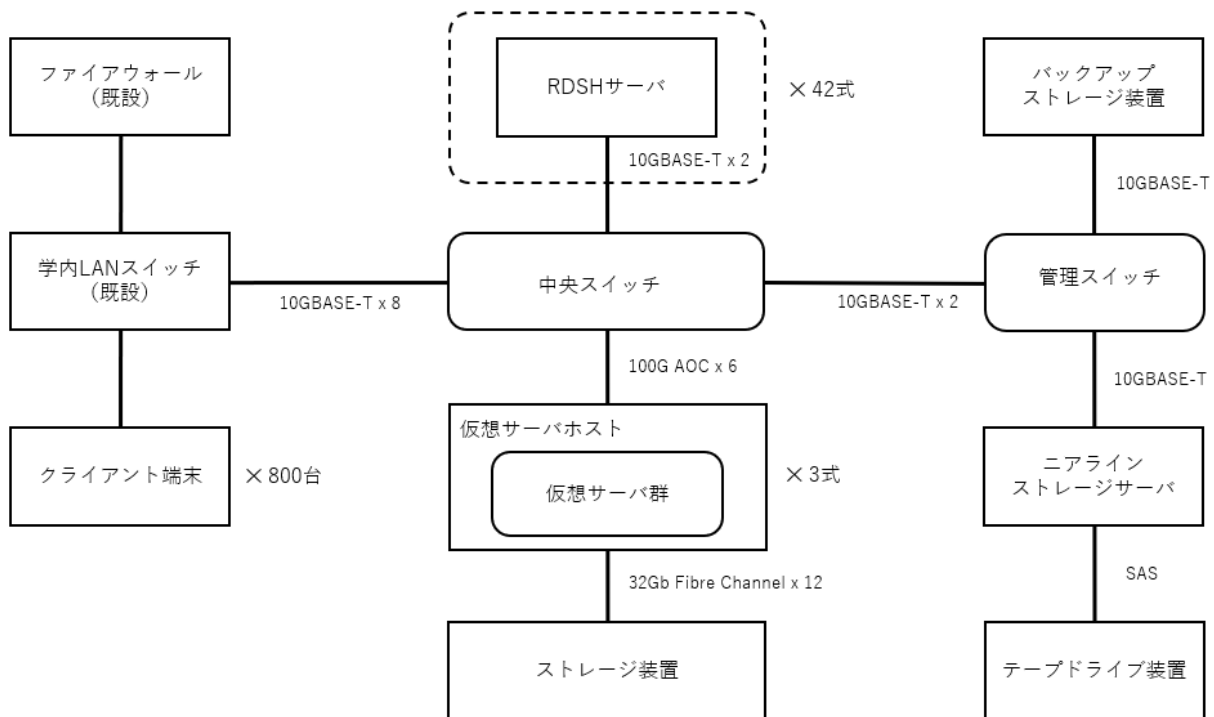


図1 システム構成図

3. アーカイブ領域の設計

テープストレージは LTF5 (Linear Tape File System) [2]を使用して管理し、テープへの読み書きはシステム管理者のみが実施することとした。そのため、テープへの書き込みは定期的に行い、テープからの読み出しは申請により対応する。

アーカイブ領域は使用頻度が低いデータファイルの保管場所として提案しているが、上記の運用ではテープに書き込むと利用者は自由に読み出すことができなくなる。共有フォルダのデータをテープストレージに保存することは、デスクにあったファイルを鍵のかかる書庫に入れるようなものである。そのため、アーカイブ領域を以下の3つのステージに分け、年に2回、6月末と12月末にデータをマイグレーションすることで、利便性の低下が段階的になるように設計した(図2)。

- アーカイブ収録
- アーカイブ閲覧
- アーカイブ閉架

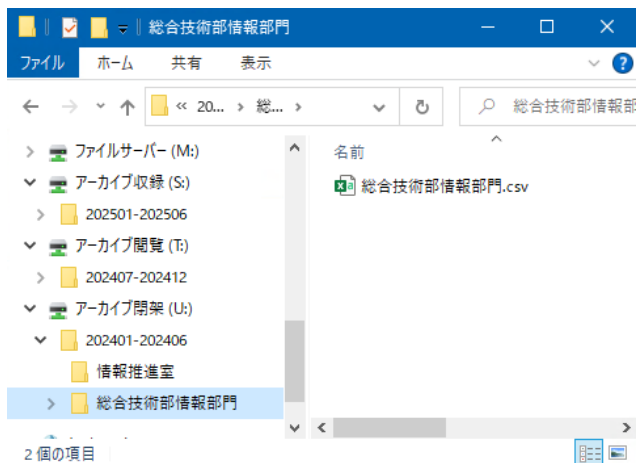


図2 アーカイブ領域

3.1 アーカイブ収録

最初のステージは「アーカイブ収録」である。この領域は共有フォルダにあるデータファイルをアーカイブ領域に書き込むための領域で、SSDストレージ上にある。収録期間を表すフォルダ(例えば、2024年後半であれば「202407-202412」)の配下に所属部署ごとに読み書き可能なフォルダを用意しており、利用者は部署フォルダの配下であればサブフォルダやファイルを作成、編集、削除可能である。

収録期間終了後、「アーカイブ収録」のデータを「アーカイブ閲覧」にマイグレーションし、「アーカイブ収録」には新しい収録用フォルダを用意する。

3.2 アーカイブ閲覧

次のステージは「アーカイブ閲覧」で、この領域もSSDストレージ上にある。アーカイブ収録にある部署フォルダのうち、データが収録されたフォルダのみをマイグレーションし、アクセス権を所属読み取り専用に変更する。アーカイブ領域へのデータ移行後も、利用者が自由に読み出せる期間を提供することで、利用者がアーカイブ領域へデータ移行するハードルを下げるのが目的である。

収録データをマイグレーションする際、この「アーカイブ閲覧」にあった1世代前の閲覧データは次の「アーカイブ閉架」にマイグレーションする。

3.3 アーカイブ閉架

最後のステージは「アーカイブ閉架」で、アーカイブ閲覧からマイグレーションされたデータはテープストレージに保存される。利用者はテープストレージには直接アクセスできないため、SSDストレージ上に「アーカイブ閉架」の領域を作成し、ファイルのパス、サイズ、作成日時、最終更新日時を一覧にまとめたファイルリストを部署フォルダ毎にCSV形式で配置する。

利用者から閉架フォルダの読み出し依頼があった場合、依頼されたファイルを含むサブフォルダ単位で、共有ファイルシステムにあるアーカイブ閉架の当該部署フォルダ直下に読み取り専用で書き戻す。書き戻したデータは定期的に共有ファイルシステムから削除する。

4. 運用

アーカイブ領域は2024年6月から試験運用と動作検証を行い、11月から本格運用を開始した。運用開始から最初のマイグレーション作業実施までの約2か月でアーカイブ収録には約980GBのデータが書き込まれた。それに合わせて、通常増えていくばかりの共有フォルダからは400GB以上の容量削減が見られた(図3)。

新しいデータも増えていくことから一概には言えないが、アーカイブ領域と共有フォルダのストレージ使用量の変化からは、アーカイブ領域へデータを移行しても共有フォルダから削除していないケースもあると考えられる。そのため、利用者へはアーカイブ領域へ移行したデータは共有フォルダからは削除し、重複データにならないように通知している。

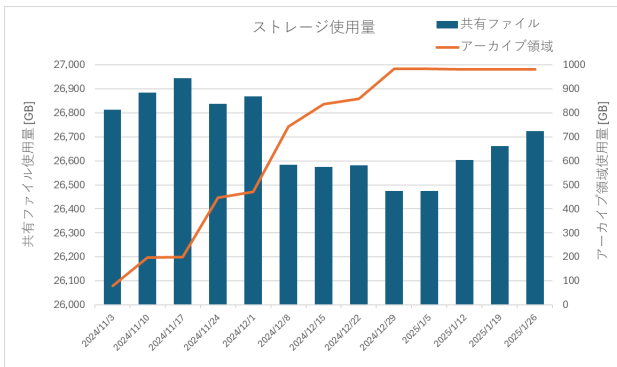


図3 ストレージ使用量

マイグレーション作業は利用者の書き込みを制限するために共有を停止して実施した。「マイグレーション収録」の共有を停止して作業を実施したが、作業中に共有フォルダにアクセスできない、事務用情報システムにログインできないなどの不具合が発生した。利用者にはグループポリシーにより当該領域をマップしているが、共有を停止したことでマップできず障害が発生したと考えられる。そのため、急遽共有を再開し、障害復旧を行った。マイグレーション作業は当該領域のトップフォルダのアクセス権限を変更して実施した。

5. 今後

事務職員用共有フォルダの有効活用のために、アーカイブ領域の導入を行った。本領域の利用により、利用頻度の低いデータをテープストレージに保存し、共有フォルダのデータ整理を進める。また、部局所有のNAS等のデータも共有ファイルシステム及びアーカイブ領域に移行することにより、安全で効率的なデータ管理を実現する。

マイグレーション作業については、利用者のアーカイブ領域へのアクセス制限方法等を改めて確認、検証し、実施手順の確立を行う。

参考文献

- [1] 浜貴幸、他、「金沢大学事務用情報システムの更改」、2023年度高エネルギー加速器研究機構技術研究会
- [2] https://ltfs.jp/about_ltfs/ltfs.html